AF/1637 #4 2FW

Docket No.: GNE.2930R1C4

September 21, 2005
Page 1 of 2

Please Direct All Correspondence to Customer Number **30313**

TRANSMITTAL LETTER

APPEAL BRIEF

Applicant	:	Botstein, et al.
App. No	:	10/033,396
Filed	:	December 27, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPYIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Fredman, J.
Art Unit	:	1637

CERTIFICATE OF MAILING

I hereby certify that this correspondence and all marked attachments are being deposited with the United States Postal Service as first-class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on

September 21, 2005
(Date)

AnneMarie Kaiser, Reg. No. 37,649

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Transmitted herewith for filing in the above-identified application are the following enclosures:

(X) Appeal Brief in 54 pages.

FILING FEES:

FEE CALCULATION				
FEE TYPE		**FEE CODE**	**CALCULATION**	**TOTAL**
Appeal Brief	*41.20(b)(2)*	1402 ($500)		$500
1 Month Extension	*1.17(a)(1)*	1251 ($120)		$
2 Month Extension	*1.17(a)(2)*	1252 ($450)		$
3 Month Extension	*1.17(a)(3)*	1253 ($1,020)		$
			TOTAL FEE DUE	$500

(X) A copy of evidence cited in Appellant's Brief and listed in Appendix B.

(X) An Amendment After Final Office in 3 pages is enclosed.

Docket No. : GNE.2930R1C4 **Customer No.: 30,313**

Application No. : 10/033,396

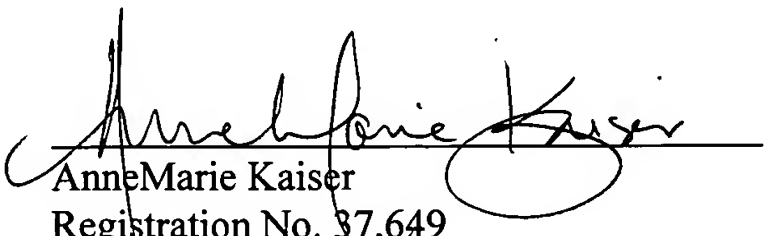
Filing Date : December 27, 2001

(X) A check in the amount of $500 is enclosed.

(X) Return prepaid postcard.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Dated: September 21, 2005



AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

1945662
092105

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Botstein, et al.
Appl. No.	:	10/033,396
Filed	:	December 27, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Fredman, J.
Group Art Unit	:	1637

CERTIFICATE OF MAILING

I hereby certify that this correspondence and all marked attachments are being deposited with the United States Postal Service as first-class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on

September 21, 2005
(Date)

AnneMarie Kaiser, Reg. No. 37,649

ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES
APPELLANT'S BRIEF

Mail Stop Appeal Brief – Patents
COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

The Applicants appeal the rejection of Claims 22-26 in the above-captioned patent application. These claims were rejected in a Final Office Action mailed April 25, 2005. Applicants filed a Notice of Appeal July 22, 2005. Applicants file concurrently herewith an Amendment After Final Office Action.

I. REAL PARTY IN INTEREST

Pursuant to 37 C.F.R. 41.37(c)(1), Appellants hereby notify the Board of Patent Appeals and Interferences that the real party in interest is the assignee of record for this application, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080.

II. RELATED APPEALS AND INTERFERENCES

A Notice of Appeal has been filed in the related Application No. 10/032996. Appellants are unaware of any other related appeals or interferences.

III. STATUS OF THE CLAIMS

The above-captioned application was filed with Claims 1-21. Claims 1-21 were canceled and new Claims 22-27 were added in a Preliminary Amendment mailed December 27, 2001. Claim 27 was canceled in an Amendment and Response to Final Office Action dated December 6, 2004. Claims 22-26 were finally rejected by the Examiner in a final Office Action mailed April 25, 2005. Appellants file concurrently herewith an Amendment After Final Office Action amending Claim 22. Accordingly, Claims 22-26 are the subject of this appeal. The claims attached hereto as Appendix A reflect the claims as amended by the Amendment After Final Office Action filed concurrently herewith.

IV. STATUS OF AMENDMENTS

Appellants mailed a "Request for Continued Examination" on March 7, 2005, submitting the claims with an amendment to Claim 22, and offering additional evidence in support of Appellants' arguments. In a Final Office Action mailed April 25, 2005, the Examiner indicated that the submission filed on March 9, 2005 had been entered. Appellants file concurrently herewith an Amendment After Final Office Action further amending Claim 22.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

The claimed subject matter relates to isolated antibodies which specifically bind to the polypeptide having SEQ ID NO:7. As amended, with the amended text shown underlined, independent Claim 22 reads:

22. An isolated antibody that specifically binds to the polypeptide of SEQ ID NO:7.

Various aspects of the claimed antibody are described in the specification at, for example, page 12, lines 8-9; page 31, lines 10-12; page 44, lines 13-18; page 46, line 25 through page 48, line 7; and page 90, line 20 through page 98, line 30. SEQ ID NO: 7 is disclosed in the Sequence Listing appended to the application.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The Examiner has rejected Claims 22-26 under 35 U.S.C. §101, stating that the claimed invention is not supported by either a specific and substantial asserted utility or a well-established utility.

The Examiner has also rejected Claims 22-26 under 35 U.S.C. §112, first paragraph. The Examiner asserts "that it would require undue experimentation for one of skill in the art to perform the method of the claim as broadly written."

VII. APPELLANTS' ARGUMENT

A. Summary of Arguments

1. *Utility Rejection*

The first issue before the Board is whether Appellants have asserted at least one "specific, substantial, and credible utility." *See* Examination Guidelines, 66 Fed. Reg. 1092 (2001). Appellants have asserted that the claimed antibodies to the polypeptide of SEQ ID NO: 7 (the PRO539 polypeptide) are useful as diagnostic tools for cancer, particularly for lung and colon cancer. This asserted utility is specific, substantial, and credible.

Briefly stated, Appellants' asserted utility is based on the disclosure in Example 16 of the instant application that the gene encoding the PRO539 polypeptide is amplified at least two-fold in the majority of the lung and colon tumors tested compared to normal tissue. It is well-established that gene amplification is correlated with an overexpression of the corresponding mRNA and the encoded polypeptide. Thus, one of skill in the art would be more likely than not to believe that the amplification of the PRO539 gene in a majority of lung and colon tumors leads to an increase in PRO539 mRNA and protein in these lung and colon tumors compared to their normal tissue counterpart. This differential expression of PRO539 mRNA and polypeptide is useful for distinguishing lung or colon tumor tissue from its normal tissue counterpart. Therefore, the claimed antibodies to the PRO539 polypeptide have a specific, substantial and credible utility as diagnostic tools for cancer, particularly lung and colon cancer, as is explained in more detail below.

2. *Enablement Rejection*

The second issue before the Board is whether Appellants have enabled the pending claims such that one of skill in the art would be able to make and use the claimed invention. The Examiner has rejected Claims 22-26 under 35 U.S.C. §112, first paragraph, arguing that the claimed subject matter was not described in the specification in such a way as to enable one

skilled in the art to make and/or use the invention. *Office Action* at 7. The Examiner recites the factors for determining enablement from *In re Wands*, 858 F.2d 731, 8 U.S.P.Q. 2d 1400 (Fed. Cir. 1988), relying to a large extent on the arguments made above in support of the rejection for lack of utility.

Appellants submit that Claims 22-26 are enabled such that one of skill in the art could make and use the claimed antibodies without undue experimentation. It is well within those of skill in the art to make antibodies which are specific to a disclosed sequence, and the Examiner has not stated otherwise. *See also In re Wands*, 858 F.2d 731 (reversing the Board's decision of non-enablement and holding that as of 1980, undue experimentation was not required to make high-affinity monoclonal antibodies to a target peptide).

Appellants assert that the claimed antibodies are useful as diagnostic tools for cancer, particularly lung and colon cancer based on the amplification of the PRO539 gene in lung and colon tumors. As detailed below, it is well-established that gene amplification is correlated with an overexpression of the encoded polypeptide, and thus it is likely that the PRO539 polypeptide is overexpressed in lung and colon cancer. Thus, based on the disclosure in the application, one of skill in the art would be able to use the claimed antibodies as diagnostic tools to detect overexpression of the PRO539 polypeptide in suspected lung and colon tumors without undue experimentation.

B. Utility Rejection –Detailed Arguments

The first issue before the Board is whether Appellants have asserted at least one "specific, substantial, and credible utility." *See* Examination Guidelines, 66 Fed. Reg. 1092 (2001). Appellants have asserted that the claimed antibodies to the polypeptide of SEQ ID NO: 7 (the PRO539 polypeptide) are useful as diagnostic tools for cancer, particularly for lung and colon cancer. This asserted utility is specific, substantial, and credible, as is explained in more detail below.

1. *Utility – Legal Standard*

A "specific utility" is defined as utility which is "specific to the subject matter claimed," in contrast to "a general utility that would be applicable to the broad class of the invention." *See*

M.P.E.P. § 2107.01 I. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic tool without also identifying the condition that is to be diagnosed.

The requirement of "substantial utility" defines a "real world" use, and derives from the Supreme Court's holding in *Brenner v. Manson*, 383 U.S. 519, 534 (1966) stating that "[t]he basic *quid pro quo* contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility." In explaining the "substantial utility" standard, M.P.E.P. § 2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase "immediate benefit to the public" or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be "currently available" to the public in order to satisfy the utility requirement. "Rather, any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a 'substantial' utility." *M.P.E.P.* § 2107.01 (emphasis added).

Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement, set forth in M.P.E.P. § 2107 II(B)(1) gives the following instruction to patent examiners: "If the applicant has asserted that the claimed invention is useful for any particular practical purpose ... and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility."

Finally, in assessing the credibility of the asserted utility, the M.P.E.P. states that "to overcome the presumption of truth that an assertion of utility by the applicant enjoys" the PTO must establish that it is "more likely than not that one of ordinary skill in the art would doubt (i.e., 'question') the truth of the statement of utility." *M.P.E.P.* § 2107.02 III A.

2. ***Utility – Burden of Proof***

It is well established that a specification which contains a disclosure of utility which corresponds in scope to the subject matter sought to be patented "must be taken as sufficient to satisfy the utility requirement of § 101 for the entire claimed subject matter unless there is reason for one skilled in the art to question the objective truth of the statement of utility or its scope." *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. 288, 297 (C.C.P.A. 1974). Thus "the PTO has the initial burden of challenging a presumptively correct assertion of utility in the disclosure." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). Only after the PTO provides

evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence sufficient to convince such a person of the invention's asserted utility. *Id.*

3. ***Utility – Standard of Proof***

Compliance with 35 U.S.C. § 101 is a question of fact. *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. 592, 596 (Fed. Cir. 1983). The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard. *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992). This is stated explicitly in the M.P.E.P.:

> [T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty. Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (emphasis in original, citations omitted).

The Court of Appeals for the Federal Circuit has stated that the standard for satisfying the utility requirement is a low one:

> The threshold of utility is not high: An invention is "useful" under section 101 if it is capable of providing some identifiable benefit. *See Brenner v. Manson,* 383 U.S. 519, 534, 86 S.Ct. 1033, 16 L.Ed.2d 69 (1966); *Brooktree Corp. v. Advanced Micro Devices, Inc.,* 977 F.2d 1555, 1571 (Fed. Cir. 1992) ("To violate § 101 the claimed device must be totally incapable of achieving a useful result"); *Fuller v. Berger,* 120 F. 274, 275 (7th Cir. 1903) (test for utility is whether invention "is incapable of serving any beneficial end"). *Juicy Whip, Inc. v. Orange Bang, Inc.*, 185 F.3d 1364, 1366, 51 U.S.P.Q. 2d 1700 (Fed. Cir. 1999) (emphasis added).

The low threshold for satisfying the utility requirement is reflected in the standard set by the Federal Circuit for invalidating a patent based on a lack of utility: "[T]he fact that an invention has only limited utility and is only operable in certain applications is not grounds for finding lack of utility. Some degree of utility is sufficient for patentability. Further, the defense of non-utility cannot be sustained without proof of total incapacity." *Envirotech Corp. v. Al George, Inc.*, 730 F.2d 753, 762, 221 U.S.P.Q. 473 (Fed. Cir. 1984) (emphasis added, citations omitted).

Because the standard for satisfying the utility requirement is so low, requiring total incapacity for a finding of no utility, the M.P.E.P. cautions that:

> Rejections under 35 U.S.C. 101 have been *rarely* sustained by federal courts. Generally speaking, in these *rare* cases, the 35 U.S.C. 101 rejection was sustained [] because the applicant ... asserted a utility that could only be true if it violated a scientific principle, such as the second law of thermodynamics, or a law of nature, or was wholly inconsistent with contemporary knowledge in the art. *M.P.E.P.* § 2107.02 III B., citing *In re Gazave*, 379 F.2d 973, 978, 154 U.S.P.Q. 92, 96 (C.C.P.A. 1967) (underline emphasis in original, italic emphasis added).

4. *Appellants Asserted a Specific, Substantial and Credible Utility that is Sufficient to Satisfy the Utility Requirement of § 101*

The claimed subject matter relates to antibodies which specifically bind to the polypeptide having SEQ ID NO: 7. The polypeptide of SEQ ID NO: 7 (referred to as "PRO539 polypeptide") is encoded by the polynucleotide of SEQ ID NO: 6 (also referred to as DNA47465-1561). *Specification* at 35, lines 20-23. Appellants have asserted that the claimed antibodies are useful as diagnostic tools for cancer, particularly lung and colon cancer.

In "Example 16: Gene Amplification" Appellants disclose that the gene encoding PRO539 polypeptide is amplified at least 2-fold in a majority of lung and colon tumors tested compared to normal tissue, respectively (a ? Ct value of 1 corresponds to a 2-fold increase, 2 ? Ct corresponds to 4-fold, 3 ? Ct corresponds to 8-fold, *etc.*). *Specification* at 111, line 35 through page 117. As explained on page 112, line 6 through page 113, line 31 of the specification, the amplification of the PRO539 gene was detected using the well-established technique of quantitative TaqMan PCR amplification of genomic DNA isolated from lung and colon tumors. *See also, Declaration of Audrey Goddard, Ph.D.*

The specification teaches that amplification of a PRO polypeptide-encoding gene in one or more tumor tissues as compared normal tissue "is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention... and diagnostic determination of the presence of those cancers." *Specification* at 112, lines 0-3. Likewise, Appellants disclose the use of antibodies to PRO polypeptides as diagnostic tools:

> [A]nti-PRO antibodies may be used in diagnostic assays for PRO, *e.g.*, detecting its expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases. *Specification* at 98, lines 6-9.

Taken together, the specification clearly discloses the use of the claimed antibodies as diagnostic tools for cancer, particularly lung and colon cancer. This utility is substantial, as one

of skill in the art will recognize that the diagnosis of cancer is a "real world" use; it is specific, as the diagnosis of lung and colon cancer is not a utility that applies to the broad class of antibodies; and it is credible, as it is not a utility "that could only be true if it violated a scientific principle, ...or a law of nature, or [is] wholly inconsistent with contemporary knowledge in the art." *M.P.E.P.* § 2107.02 III B., citing *In re Gazave*, 379 F.2d 973, 978, 154 U.S.P.Q. 92, 96 (C.C.P.A. 1967).

Because Appellants' specification contains a disclosure of utility which corresponds in scope to the claimed subject matter, antibodies which specifically bind to the polypeptide of SEQ ID NO: 7, the asserted utility "must be taken as sufficient to satisfy the utility requirement of § 101 for the entire claimed subject matter unless there is reason for one skilled in the art to question the objective truth of the statement of utility or its scope." *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. 288, 297 (C.C.P.A. 1974). Therefore, the burden of establishing a *prima facie* case of lack of utility rests with the PTO. *See, In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995) ("the PTO has the initial burden of challenging a presumptively correct assertion of utility in the disclosure").

5. ***The Examiner's Arguments***

To establish a *prima facie* showing that the claimed subject matter lacks utility, the Examiner must "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The Examiner has filed a first Office Action on April 19, 2004, a final Office Action on September 8, 2004, an Advisory Action on December 20, 2004, and a second final Office Action on April 25, 2005 after Appellants' Request for Continued Examination. None of these papers provide sufficient evidence that one of ordinary skill in the art would reasonably doubt the asserted utility.

The final Office Action on April 25, 2005, contains all of the Examiner's arguments in support of the utility rejection made in the previous office actions. Appellants summarize these arguments below.

a. ***Appellants' evidence is allegedly deficient***

The Examiner has asserted that that the evidence of PRO539 gene amplification is deficient, stating:

> The level of overexpression of PRO 539 was minimal, not even a two fold overexpression in any cell type. Further, no statistical data was presented to show that the overexpression was significant in any way with no P-value or other statistical measure to demonstrate that. the overexpression was a real effect and not simply produced by chance. *Office Action* at 3[1].

b. ***The Examiner asserts that there is no "necessary" connection between DNA amplification, mRNA level, and protein level***

The critical argument for the Examiner's rejection is that "the overexpression of the nucleic acid is not relevant to the utility of the protein and antibody." *Office Action* at 3. The Examiner cites Meric *et al.* (Mol. Cancer Ther. (2002) 1:971-979) Gökman -Polar *et al.* (Cancer Res. (2001) 61:1375-1381) and Pennica *et al.* (Proc. Natl. Acad. Sci. USA (1998) 95:14717-14722) to support his assertion that "not only is there no necessary connection between the level of protein in a cell and the amount of mRNA, but there is also no necessary correlation between the amount of DNA in a cell and the amount of mRNA." *Office Action* at 4 (emphasis added). Therefore, the Examiner concludes that "any evidence by Applicant showing overexpression of one component does not provide utility for the protein itself." *Id.* at 4-5.

c. ***The Examiner asserts that a protein's function cannot be predicted from its sequence***

The Examiner also argues that "given the breadth of these claims which encompass any antibody to the Pro-539 polypeptide, there is an abundance of evidence that very similar proteins can perform very different functions." *Office Action* at 5. The Examiner cites Rost *et al.* (J. Mol. Biol. (2002) 318(2):595-608) to support the assertion that "even high levels of homology do not necessarily correlate with actual protein function. In the current case, where the function of PRO-539 (SEQ ID NO: 7) is not known, the expectation is even lower that there is any utility that can be derived based upon the sequence." *Id.*

---

[1] Citation to "Office Action" refers to the final Office Action mailed April 25, 2005 unless otherwise indicated.

d. ***The Examiner asserts that the instant case is similar to Examples 4 and 12 of the Utility Guidelines***

The Examiner argues that the instant case is extremely similar to Example 12 of the Utility Guidelines where a protein which was known to be a receptor, but where the ligand was unknown, was found to lack utility. The Examiner argues that "[i]n the current case, the putative PRO-539 protein, lacks any substantial utility whatsoever, and solely relies upon an small level of mRNA overexpression in cancer cells. However, there is no necessary relationship between the protein levels or utilities and such an overexpression of the nucleic acid." *Office Action* at 5 (emphasis added). The Examiner concludes that further research would be required to identify and reasonably confirm a "real world" context of use for PRO-539 antibodies. Without explanation, the Examiner also asserts that the instant case "directly tracks" Example 4 of the Utility Guidelines as well. *Office Action* at 6.

e. ***The Examiner asserts that the claimed invention lacks a specific utility***

In addition to the arguments above that the claimed invention lacks a substantial utility, the Examiner also asserts that the claimed invention lacks a specific utility. The Examiner argues that "the antibody to the protein, as distinguished from the nucleic acid, has not been associated with any disease, any condition, or any other specific feature. There is no association of the antibody or protein with cancer or with any other disease." *Office Action* at 6. The Examiner states that "a general statement of diagnostic utility, such as diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed." *Id.*

6. ***The Examiner has not established a prima facie case that Claims 22-26 lack Utility***

The above arguments do not satisfy the Examiner's burden to "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The Examiner has the burden of presenting "countervailing facts and reasoning sufficient to establish that a person of ordinary skill would not believe the applicant 's assertion of utility." *M.P.E.P.* §2107.02 III.A., *citing in*

re Brana, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995) ("Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence"). The Examiner's argument hinges on his assertion that there is no "necessary connection" or "necessary correlation" between gene amplification and mRNA or protein overexpression. The Examiner cites three references to support this assertion.

The Examiner's arguments fail for two reasons. First, the lack of a "necessary," *i.e.*, precise or perfect, connection is not sufficient to establish that one of skill in the art would reasonably doubt Appellants' asserted utility. Appellants do not need to establish the utility "beyond a reasonable doubt" or to a "statistical certainty." Instead, Appellants need show that one of ordinary skill in the art would conclude that the asserted utility is more likely than not true. *See M.P.E.P.* § 2107.02, part VII. Thus, the lack of a "necessary" connection does not mean that Appellants' asserted utility is not true under a more likely than not standard. For example, if the presence of A results in the presence of B in 9 out of 10 cases there is no "necessary" connection between A and B, yet one would believe that it is more likely than not true that given A, B will be present.

The second reason the Examiner's arguments fail is that the references relied on by the Examiner are not contrary to Appellants' asserted utility, and in fact actually support Appellants' position. Because the references relied on by the Examiner offer little or no support for his assertions, the Examiner has failed to establish a *prima facie* showing that one of skill in the art would reasonably doubt the asserted utility, and the Board should accept Appellants' disclosed utility as sufficient. Appellants address each of the Examiner's arguments in turn.

a. The data in Example 16 are sufficient to establish the asserted utility

Appellants turn first to the Examiner's arguments challenging the reliability of the data reported in Example 16. The Examiner makes two arguments regarding the sufficiency of the data in Example 16: (1) he argues that the level of overexpression of PRO539 was minimal, "not even a two fold overexpression in any cell type;" and (2) that there is no statistical analysis to show that the data were significant. *Office Action* at 3. Appellants submit that the gene amplification data provided in Example 16 are sufficient to establish a specific and substantial utility for the PRO539 polypeptide and antibody.

As described in Example 16 of the present application, the inventors isolated genomic DNA from a variety of primary cancers and cancer cell lines. As a negative control, DNA was isolated from the blood of normal healthy individuals. *Specification* at 112, lines 7-9. Gene amplification was monitored using real-time quantitative TaqMan™ PCR. The gene amplification results are set forth in Table 7 (Table 8 as amended) on page 117. As explained in the specification on page 112, lines 17-19, the results of TaqMan™ PCR are reported in ΔCt units. It is well-known in the art that "Ct" stands for "threshold cycle." One Ct unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, 2 units correspond to 4-fold amplification, 3 units to 8-fold amplification, *etc.* *Specification* at 112, lines 17-19.

Looking at the results reported in on page 117, nine primary lung tumors and eight primary colon tumors were tested, as well as a number of tumor cell lines. PRO539 had a ΔCt value of greater than 1, *i.e.*, more than two-fold amplification, in six of nine lung tumors and five of eight colon tumors. These data show that in more than half of the lung and colon tumors tested, the gene for PRO539 was amplified at least two-fold. Thus, the Examiner's statement that the overexpression of PRO539 was minimal, "not even a two fold overexpression in any cell type" is simply wrong.

The Examiner has also rejected the data because he questions the statistical significance of the data. However, Appellants are not required to prove utility to a statistical certainty, only that it is more likely than not true. *See Nelson v. Bowler*, 626 F.2d 853, 856-57, 206 U.S.P.Q. 881, 883-84 (C.C.P.A. 1980) (reversing the Board and rejecting an argument that evidence of utility was insufficient because it was not statistically significant). As the M.P.E.P. states:

> [T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." **Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty**. Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (bold emphasis added, underline in original, citations omitted).

Therefore, whether the results are statistically significant or not is irrelevant to establishing the asserted utility. The results must simply be reliable enough that one of skill in the art would believe that the utility is more likely than not true.

In support of Appellants' assertion of utility, Appellants submitted a copy of the declaration of Dr. Audrey Goddard with exhibits A-G (the Goddard Declaration). As Dr.

Goddard's *curriculum vitae*, Exhibit A of the Goddard Declaration, shows, she is an expert in the art of identifying and quantifying the amplification of oncogenes in cancers.

In her declaration, Dr. Goddard states that

> [T]he quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that ... a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy. *Goddard Declaration* at ¶ 7 (emphasis added).

Therefore, according to Dr. Goddard, an expert in the field, a 2-fold or more increase, *i.e.*, a ΔCt value of 1 or greater, is significant and reliable enough to be useful for the diagnosis of cancer. The ΔCt value for PRO539 is greater than 1 for the majority of primary lung and colon tumors tested, indicating a more than two-fold amplification. *Specification* at page 117.

The Examiner has not offered any substantial reason or evidence to reject either the underlying data or Dr. Goddard's conclusions. Instead, the Examiner has dismissed the Goddard declaration as "fundamentally flawed," arguing the it fails to provide specific evidence regarding PRO539. *Final Office Action* mailed 9/8/04 at 11. While it is true that Dr. Goddard's Declaration discusses the significance of the data generically without specific reference to PRO539, her declaration read in light of the data reported in Example 16 clearly make her declaration applicable to PRO539.

Dr. Goddard is an expert in the field whose declaration is based on personal knowledge of the relevant facts at issue. Appellants' have reminded the Examiner that "Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned." *M.P.E.P.* § 2107 (emphasis added). In addition, declarations relating to issues of fact should not be summarily dismissed as "opinions" without an adequate explanation of how the declaration fails to rebut the Examiner's position. *See in re Alton* 76 F.3d 1168 (Fed. Cir. 1996). Therefore, the Examiner should accept Dr. Goddard's statement that "a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer."

Interestingly, the above argument along with the Goddard Declaration were presented to Examiner Fredman in the closely related and co-owned patent application Serial No. 10/033,167 to overcome a nearly identical 35 U.S.C. § 101 utility rejection of claims directed in part to the

nucleic acid sequence encoding the PRO539 protein. That application contains the same data presented and relied on in the instant application. In that case, the Examiner concluded that the Goddard declaration was "sufficient to overcome the rejection of the claims based upon utility and enablement as discussed above." *Office Action* mailed 9/9/2003, page 5.

Thus, it is the uncontested opinion of an expert in the field that the results in Example 16 are significant and reliable enough to indicate that the PRO539 gene is useful as a diagnostic tool for cancer, and the PTO has accepted this utility in a related application.

In conclusion, Appellants submit that the evidence reported in Example 16, supported by the Goddard Declaration, establish that the PRO539 gene is amplified at least two-fold in the majority of lung and colon tumors tested and can be used as a marker for the diagnosis of cancer. The Examiner has not offered any significant arguments or evidence to the contrary, and therefore has not established a *prima facie* case that one of skill in the art would reasonably doubt the significance of the data in Example 16, or the assertion that the PRO539 gene can be used as a diagnostic tool. For the reasons discussed below, based on these data Appellants' have established a specific and substantial utility for the PRO539 polypeptide and antibody as well.

b. The Examiner's cited references are not contrary to Appellants' assertion that gene amplification generally leads to increased mRNA and protein levels

Appellants next address the heart of the Examiner's rejection – his assertion that "not only is there no necessary connection between the level of protein in a cell and the amount of mRNA, but there is also no necessary correlation between the amount of DNA in a cell and the amount of mRNA. Therefore, any evidence by Applicant showing overexpression of one component does not provide utility for the protein itself." *Office Action* at 4-5 (emphasis added).

As an initial matter, Appellants emphasize that the level of proof required to establish utility is relatively low. "[T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true 'beyond a reasonable doubt.' Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty." *M.P.E.P.* § 2107.02, part VII (citations omitted). Therefore, there does not have to be a "necessary," precise or perfect correlation between gene amplification and increased mRNA and

protein levels. Instead, the evidence must be sufficient to convince a skilled artisan that the asserted utility is "more likely than not true." *Id.*

Thus, even if the references cited by the Examiner supported the conclusion that there is no "necessary" connection between gene amplification, mRNA, and protein levels, this is not the requisite standard which Appellants must meet to satisfy their burden. Instead, there need only be evidence that the asserted utility is "more likely than not true." As discussed above, it does not follow that one of skill in the art would not believe that the asserted utility is more likely than not true *even if* the connection between gene amplification and the level of mRNA and protein is not "necessary." Stated another way, the correlation between gene amplification and the level of mRNA and protein does not have to be exact for the asserted utility to be more likely than not true.

Keeping the standard of proof in mind, we turn to the three references the Examiner cites as support for his assertion that there is no "necessary" connection between gene amplification, the amount of mRNA and the amount of protein. The first is an article by Meric *et al.* The Examiner states that "in a discussion of regulation of gene activity in cancer [Meric] notes that 'Gene expression is quite complicated, however, and is also regulated at the level of mRNA stability, mRNA translation and protein stability (page 971, column 1).'" *Office Action* at 4. From this statement, the Examiner concludes that "Meric teaches that there is not necessarily a correlation between mRNA levels and protein levels in cancer cells, since the regulation may occur at levels other than that of the mRNA, such as in the level of translation of the mRNA or in the stability of the protein." *Office Action* at 3-4.

What the Examiner ignores is the preceding statement by the authors:

> The **fundamental principle** of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells...[M]ost efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable to either DNA amplification or to differences in transcription. Meric at 971 (emphasis added).

This statement supports Appellants' asserted utility. First, Meric teaches that although there is more than one mechanism, changes in mRNA in tumors cells can be attributed to gene amplification. This clearly supports Appellants' assertion that gene amplification leads to increase in mRNA level.

Second, Meric states that examining changes in mRNA levels is a "fundamental principle" of molecular cancer therapeutics. It is true that there is no necessary, *i.e.*, absolute or perfect, correlation between gene expression and protein expression because there are other mechanism for regulating gene expression. However, exploiting differences in gene expression between cancer cells and normal cells would not be a "fundamental principle" of molecular cancer therapeutics if there were no significant correlation between gene expression and protein levels. Stated another way, changes in mRNA without corresponding changes in protein levels would have little or no effect on cellular biology, and those of skill in the art would have no reason to examine the differences in gene expression at the mRNA level without such a correlation. However, as one of skill in the art recognizes, there is a strong correlation between changes in mRNA and changes in protein level. It is because of this strong correlation that it remains a "fundamental principle" of molecular therapeutics in cancer to look at changes in mRNA level.

Thus, while Meric does teach that there is no "necessary" correlation between changes in mRNA and changes in protein, it supports Appellants' assertion that generally there is a good correlation between the two. Appellants reiterate that they need not prove an asserted utility is true "beyond a reasonable doubt" or as a matter of "statistical certainty." The evidence must lead a person of ordinary skill in the art to conclude that the asserted utility is "more likely than not true." *M.P.E.P.* § 2107.02, part VII. Because the correlation does not have to be absolute for one of skill in the art to believe the asserted utility is more likely than not true, Meric does not support the Examiner's rejection, and does not satisfy his burden of presenting "countervailing facts and reasoning sufficient to establish that a person of ordinary skill would not believe the applicant's assertion of utility." *M.P.E.P.* §2107.02 III.A., *citing in re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995).

The Examiner's second citation is an article by Gökman-Polar. The Examiner asserts that the article "teaches 'Quantitative reverse transcription-PCR analysis revealed that PKC mRNA levels do not directly correlate with PKC protein levels, indicating that PKC isozyme expression is likely regulated at the posttranscriptional/translational level (see abstract).'" *Office Action* at 4. The Examiner states that Gökman -Polar shows in figures 6 and 7 that "there is no increase in mRNA expression for any of the isozymes, while the protein is significantly

overexpressed as shown by figures 4 and 5. This demonstrates that there is no relationship between mRNA levels and protein levels." *Office Action* at 4 (emphasis added).

Appellants submit that like Meric, Gökman -Polar is not contrary to Appellants' assertion that gene amplification generally leads to overexpression of the mRNA and protein. Appellants emphasize that they are asserting that gene amplification leads to an increase in PRO539 mRNA, which results in increased PRO539 protein in the tumors where PRO539 is amplified. Appellants are not asserting that every change in protein level is a result of gene amplification or an increase in mRNA expression – Appellants acknowledge that there are other ways to increase protein levels. Therefore, the lack of an large increase in PKC ßII mRNA accompanying an increase in PKC ßII protein levels, cited by the Examiner as support for his rejection, is not very relevant to Appellants' asserted utility since it does not establish that increased mRNA does not lead to increased protein. Instead, it shows that not all protein increases are a result of increased mRNA.

In addition, a close review of the entire Gökman -Polar article indicates that the trend in the data is that changes in mRNA levels are positively correlated with protein levels, supporting Appellants' assertion that the two are generally correlated. In Figure 6, the mRNA levels of two isozymes shows a decrease, while the third is increased slightly. This same pattern is seen for the corresponding protein levels in Figure 2, although as the Examiner points out, the increase in mRNA for the third isozyme is small compared to the increase in protein level. Similarly, comparing the mRNA levels of the three isozymes in Figure 7 to the corresponding protein levels in Figure 4, with one exception the changes in mRNA levels are positively correlated to changes in protein levels. While protein levels do not increase or decrease in direct proportion to the changes in mRNA, the trend in five of the six examples is that changes in mRNA levels are positively correlated to changes in protein levels.

This evidence is hardly sufficient to establish that one of skill in the art would reasonably doubt that there is a reasonable correlation between changes in mRNA levels and changes in the corresponding protein levels. First, it is evidence taken from only two genes, one of which has two splice variants. This is hardly sufficient evidence to refute Applicants' assertion that in general, increases in mRNA lead to increases in the corresponding protein. Second, even within this small sample, five of the six data points show a positive correlation between changes in mRNA and protein levels – decreases in mRNA are associated with decreases in protein, while

increases in mRNA are associated with increases in protein. Therefore, Gökman -Polar does not satisfy the Examiner's burden to establish that one of skill in the art would reasonably doubt the asserted utility, and it definitely does not support the Examiner's obviously false statement that "there is no relationship between mRNA levels and protein levels." *Office Action* at 4 (emphasis added).

Finally, the Examiner cites Pennica *et al.*, stating that the article "shows that WISP-2 DNA was amplified in cancer cells but was actually demonstrated REDUCED RNA expression (see abstract)." *Office Action* at 4 (emphasis in original). Based on this statement, the Examiner concludes that "[t]his provides additional evidence that there is no relationship between gene amplification and mRNA levels, since mRNA levels have no necessary correlation with gene amplification." *Id.* (emphasis added).

The Examiner's conclusion that there is no relationship between gene amplification and mRNA levels is not supported by the reference. Pennica reports that:

> An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP*-3 RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient." *Pennica* at 14722, col. 1 (emphasis added).

Importantly, the authors report that the *WISP-1* gene was amplified approximately 2-fold, similar to the gene for PRO539 reported in Example 16. *Id.* at 14720, col. 2.

Appellants first note that overexpression of *WISP*-3 RNA seen in the absence of DNA amplification is not relevant to the Appellants' asserted utility – Appellants are not asserting that amplification is the only source of gene overexpression. This leaves the results reported for *WISP-1*, where gene amplification (approximately two-fold) is correlated with overexpression of the mRNA, and *WISP-2*, where gene amplification was apparently not correlated with gene overexpression.

The authors of Pennica offer an explanation for what they obviously viewed as an anomalous result for *WISP-2*: "Because the center of the 20q13 amplicon [of which *WISP-2* is a part] has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon." *Pennica* at 14722, col. 1 (emphasis added). Thus, the only example of a lack of positive correlation between gene amplification and

RNA overexpression relied on by the Examiner may be an artifact. The fact that the authors attempt to explain this anomaly only supports Appellants' argument that the accepted understanding in the art is that there is a positive correlation between gene amplification and overexpression of the mRNA.

As stated above, the standard for utility is not absolute or even statistical certainty, but rather whether one of skill in the art would believe that the asserted utility is more likely than not true. One apparent contrary example, when combined with the positive example reported in the same reference, is not sufficient to prove that a person of skill in the art would have a reasonable doubt that gene amplification is not generally correlated with increased gene expression. The Examiner has not shown whether the apparent lack of correlation observed for one of the two amplified genes studied is typical, or is merely an exception to the rule of correlation. Indeed, the authors' attempt to explain the lack of correlation suggests that this result is the exception, and not the rule.

In summary, Pennica provides one example of a positive correlation between gene amplification and increased gene expression, and one possible example of a lack of correlation. This single reference reporting a single gene that may show a lack of correlation between gene amplification and mRNA overexpression cannot support the Examiner's conclusion that there is "no relationship between gene amplification and mRNA levels" – the results of *WISP-1* clearly show that in some cases (at least half of those relied on by the Examiner) there is a relationship. For the same reason, Appellants submit that Pennica is not sufficient to satisfy the Examiner's initial burden of establishing that it is more likely than not that one of skill in the art would doubt Appellants' assertion that generally, gene amplification leads to overexpression of the mRNA and protein.

Taken together, the three references relied on by the Examiner are not sufficient to support his conclusion that "the overexpression of the nucleic acid is not relevant to the utility of the protein and antibody." *Office Action* at 3. First, Meric supports Appellants' assertion that generally, changes in mRNA lead to a corresponding change in the level of the encoded protein – that is why examining differences between tumor and normal tissue at the mRNA level is a "fundamental principle" of molecular cancer therapeutics. Likewise, Meric teaches that mRNA overexpression can be attributed to gene amplification. Second, Gökman-Polar reports a positive correlation between changes in mRNA level and changes in protein level for five of six samples

tested. Far from supporting the Examiner's conclusion that there is "no relationship" between mRNA and protein, this evidence tends to support Appellants' assertion that changes in mRNA and protein are correlated. Finally, the only reference which looked at a correlation between gene amplification and gene expression reports one gene where there was a strong correlation between the two, and one possible example where there was a lack of positive correlation. This evidence is at best inconclusive, with at least half the genes showing a correlation between gene amplification and mRNA overexpression.

While these references may establish that there is no "necessary" correlation between gene amplification and overexpression of mRNA and protein, Appellants are not required to establish the asserted utility beyond a reasonable doubt or with a statistical certainty. Thus, even assuming that the Examiner has proved that there is no "necessary" correlation, this does not mean that he has met his initial burden of establishing that it is "more likely than not" that a skilled artisan would doubt the asserted utility. Given that two of the cited references actually tend to support the Appellants' position, Appellants assert that the Examiner has failed to meet his burden of establishing a *prima facie* case of lack of utility.

c. The Examiner's arguments regarding the unpredictability of protein function from protein sequence are irrelevant

Appellants next address the Examiner's argument that "there is an abundance of evidence that very similar proteins can perform very different functions" and that "even high levels of homology do not necessarily correlate with actual protein function. In the current case, where the function of PRO-539 (SEQ ID NO: 7) is not known, the expectation is even lower that there is any utility that can be derived based upon the sequence." *Office Action* at 5.

Appellants' asserted utility does not rely on the function of the PRO539 protein, or any relation between the function of PRO539 and its sequence. The claimed subject matter relates to antibodies which specifically bind to the polypeptide of SEQ ID NO:7. The Appellants' asserted utility is the use of the claimed antibodies as diagnostic tools for cancer, based on the two-fold amplification of the PRO539 gene in a majority of the lung and colon tumors tested. This utility in no way depends on the function of the PRO539 protein, making the Examiner's arguments irrelevant.

d. The instant case is significantly different from Examples 4 and 12 of the Utility Guidelines

The Examiner argues that the instant case is extremely similar to Example 12 of the Utility Guidelines where a protein which was known to be a receptor, but where the ligand was unknown, was found to lack utility. The Examiner argues that "[i]n the current case, the putative PRO-539 protein, lacks any substantial utility whatsoever, and solely relies upon an small level of mRNA overexpression in cancer cells. However, there is no necessary relationship between the protein levels or utilities and such an overexpression of the nucleic acid." *Office Action* at 5 (emphasis added). The Examiner also asserts that the instant case "directly tracks" Example 4 of the Utility Guidelines as well. *Office Action* at 6.

First, as stated above, Appellants are not required to establish their asserted utility "beyond a reasonable doubt" or to a "statistical certainty." Therefore, any lack of a "necessary" relationship between protein levels and overexpression of the nucleic acid does not establish that one of skill in the art is more likely than not to doubt the asserted utility. Second, for the reasons stated above, the Examiner's references do not establish that one of skill in the art would doubt Appellants' assertion that gene amplification generally leads to overexpression of the mRNA and protein. For this reason, the instant case is significantly different from Examples 4 and 12 of the Utility Guidelines.

In Example 4, nothing more than the sequence of the protein was disclosed. There was no disclosed utility, or description of the chemical, physical or biological properties of the protein other than its sequence. Similarly, in Example 12, the specification discloses that an isolated protein is a receptor for an uncharacterized protein. The isolated protein is not characterized with respect to its biological function or any disease or body condition that is associated with the protein.

In contrast to Examples 4 and 12, in the instant case the specification teaches that the gene encoding the PRO539 protein is amplified at least two-fold in the majority of lung and colon tumors tested. Appellants have asserted that one of skill in the art would recognize that it is likely that the PRO539 protein is therefore also overexpressed in a majority of lung and colon tumors, an assertion supported by expert declarations and references discussed below. This biological activity distinguishes the instant case from Examples 4 and 12.

Instead, the instant case is more like the caveat to Example 12, where the additional information that the isolated protein is found on the cell membranes of melanoma cells, but not normal skin cells is disclosed. The Utility Guidelines state that this additional information is sufficient to establish utility for antibodies to the protein as diagnostic tools. Similarly, the likely overexpression of PRO539 protein in lung and colon tumors is also sufficient to establish utility for the protein and antibodies.

e. The asserted utility is specific

Finally, Appellants address the Examiner's argument that "the antibody to the protein, as distinguished from the nucleic acid, has not been associated with any disease, any condition, or any other specific feature. There is no association of the antibody or protein with cancer or with any other disease." *Office Action* at 6. The Examiner states that "a general statement of diagnostic utility, such as diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed." *Id.*

Specific Utility is defined as utility which is "specific to the subject matter claimed," in contrast to "a general utility that would be applicable to the broad class of the invention." *M.P.E.P.* § 2107.01, part I (2004). Appellants submit that the evidence of amplification and overexpression of PRO539 nucleic acids (see below) in lung and colon tumors along with the declarations and references discussed below provide a specific utility for the claimed antibodies. As discussed in more detail below, Appellants have established a reasonable correlation between gene amplification and overexpression of the mRNA and protein. The Examiner has not provided any substantial evidence or arguments to the contrary. This makes antibodies to the PRO539 protein useful in diagnosing lung and colon cancer. This is not a "a general statement of diagnostic utility" that would apply to the broad class of antibodies.

f. Conclusion – Examiner has failed to establish a prima facie case that one of skill in the art would doubt Appellants' asserted utility

The Examiner has relied on essentially two unsupported arguments in rejecting the pending claims for lack of utility. First, the Examiner has questioned the sufficiency, reliability and significance of the data reported in Example 16. Second, the Examiner has argued that because there is no "necessary" connection between gene amplification and mRNA levels, and

no "necessary" connection between mRNA levels and protein levels, evidence of gene amplification does not provide utility for the protein. Appellants have responded to each of these arguments in turn.

First, Appellants have established that the PRO539 gene is amplified at least two-fold in a majority of lung and colon tumors tested. The Goddard declaration establishes that the data in Example 16 are sufficiently reliable and significant to show that the PRO539 gene is useful as a cancer diagnostic tool. The Examiner has not provided any substantial reason or evidence for one of skill in the art to doubt the reliability or usefulness of the data in Example 16, or the facts and conclusions in the Goddard declaration.

Second, Appellants have shown that the three references relied on by the Examiner as support for his assertion that gene amplification is not relevant to the utility of the protein and antibody do not support the Examiner's position, are not contrary to Appellants' assertion of a general correlation between gene amplification and mRNA and protein overexpression, and in fact lend support to Appellants' position.

Finally, Appellants have adequately addressed the Examiners' remaining arguments regarding a lack of structure function-relationships in proteins; the similarity of the instant case to Examples 4 and 12 of the Utility Guidelines, and the asserted lack of a specific utility.

Taken together, the Examiner's arguments are not sufficient to satisfy the Examiner's burden to "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The Examiner's arguments are largely conclusory statements which are not supported by any substantial evidence or reasoning which explains why one of ordinary skill in the art would reasonably doubt the asserted utility. And even the scant evidence that is provided supports the Appellants' position as much or more than the Examiner's position. Therefore, the Board should accept the Appellants' disclosure of utility. *See Ex parte Rubin*, 5 U.S.P.Q. 2d 1461 (Bd. Pat. App. & Interf. 1987) ("There is no factual support in this record for the examiner's questioning of the denaturation test reported in the specification. ... No reason to doubt 'the objective truth' of the asserted utility having been advanced by the examiner, we accept appellant's disclosure of utility corresponding in scope to the claimed subject matter.").

7. ***Appellants have provided Sufficient Rebuttal Evidence of Utility***

"Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence." I*n re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The rebuttal evidence must be sufficient such that when it is considered as a whole, it is more likely than not that the asserted utility is true. *See In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992) (stating that the evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard). The M.P.E.P. summarizes the standard of proof required:

> [T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty. Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (emphasis in original, citations omitted).

Appellants remind the Board that the Federal Circuit has stated that the standard for satisfying the utility requirement is a low one: "The threshold of utility is not high: An invention is 'useful' under section 101 if it is capable of providing some identifiable benefit." *Juicy Whip, Inc. v. Orange Bang, Inc.*, 185 F.3d 1364, 1366, 51 U.S.P.Q. 2d 1700 (Fed. Cir. 1999).

Even if the Examiner has satisfied his burden of presenting a *prima facie* case of lack of utility, Appellants have supplied more than enough rebuttal evidence, such that when considered as a whole, one of skill in the art would conclude that the asserted utility is more likely than not true. As discussed in detail below, Appellants have provided sufficient evidence that the gene encoding the PRO539 polypeptide is amplified and overexpressed in lung and colon tumors and can therefore be used as a diagnostic tool. In addition, Appellants have shown that it is well established in the art that there is a reasonable correlation between gene amplification, changes in mRNA level and changes in the corresponding protein level such that one of skill in the art would believe that the PRO539 polypeptide is also overexpressed in lung and colon cancers. Therefore, considering the evidence as a whole, one of skill in the art would believe that it is more likely than not that the claimed antibodies are useful as diagnostic tools for cancer, particularly lung and colon tumors.

a. ***Appellants have established that the PRO539 gene is amplified at least two-fold in the majority of lung and colon tumors tested***

As discussed above, the Examiner has not provided any evidence to challenge the reliability and significance of the data in Example 16 which reports that the PRO539 gene is amplified at least two-fold in a majority of lung and colon tumors tested compared to normal tissue. In contrast to this complete lack of evidence on the part of the Examiner, Appellants have submitted the Goddard Declaration. That declaration establishes that it is the opinion of an expert in the field who has personal knowledge of the facts surrounding Example 16 that a two-fold or greater amplification of a gene in tumors makes the gene useful as a diagnostic tool. The Examiner has not provided any evidence to challenge the facts and conclusions of the Goddard Declaration in support of Example 16.

Given the disclosure of Example 16 and the supporting Goddard Declaration on the one hand, and the complete lack of any evidence or arguments on the other, it is clear that considering the evidence as a whole, one of skill in the art would conclude that it is more likely than not that the PRO539 gene is amplified at least two-fold in the majority of lung and colon tumors tested compared to normal tissue, and therefore the PRO539 gene is useful as a diagnostic tool for lung and colon cancer.

As Appellants explain below, it is more likely than not that the PRO539 mRNA and polypeptide are overexpressed in lung and colon tumors, and can therefore be used to distinguish tumor tissue from normal tissue. This provides utility for the claimed antibodies to the PRO539 polypeptide.

b. ***Appellants have established that generally there is a correlation between gene amplification and mRNA overexpression***

Appellants turn next to the evidence offered in support of their assertion that it is well-established that gene amplification generally leads to overexpression of the corresponding mRNA. The teachings in Genes V, a leading textbook in the field, illustrate that at the time the instant application was filed, it was well known by those of skill in the art that gene amplification leads to overexpression of the corresponding gene product. *See Benjamin Lewin*, Genes V, 5th ed. 1994, pages 1196-1201. In a section entitled "Insertion, translocation, or amplification may activate proto-oncogenes", the text describes various molecular events that lead to

overexpression of a gene product, using the *c-myc* gene as an example. The first mechanism taught is insertion of a retrovirus upstream of the gene which causes it to be driven by a more efficient promoter, resulting in increased mRNA and protein levels. Next, Lewin teaches that chromosomal translocations may bring a gene to a new region where it is actively expressed, resulting in increased gene and protein expression. The third mechanism whereby protein levels of oncogenes are overexpressed is gene amplification. The text emphasizes that the common thread among the different means of activation of proto-oncogenes is that the expression of the gene is increased. Thus, as of 1994, it was well-known in the art that gene amplification is correlated with overexpression of the corresponding mRNA and encoded protein.

Additional information regarding the understanding of those of skill in the art regarding the relationship between gene amplification and protein overexpression at the time the instant application was filed is found in Alitalo (Med. Biol., (1984) 62:304-317), and Merlino *et al.* (J. Clin. Invest., (1985) 75:1077-1079). Under the heading "Enhanced Expression of Amplified Oncogenes," Alitalo states that "[i]n all cases where they have been studied, the amplified oncogenes have been found abundantly expressed at the mRNA level, roughly in proportion to the amount of DNA amplification (see Table 1)." *Alitalo* at 313 (emphasis added). Table 1 lists eleven examples of amplified oncogenes where expression levels were examined. In all eleven cases, expression of the amplified oncogene was elevated. Thus, Alitalo clearly teaches that amplification leads to overexpression. Merlino *et al.* studied epidermoid carcinoma cells, and teach that amplification of the EGF receptor gene results in increased levels of EGF receptor mRNA and increased levels of EGF receptor protein. Taken together, the excerpt from Genes V, as well as the Alitalo and Merlino references, establish that as of the filing date of the instant application, those of skill in the art appreciated the correlation between gene amplification and overexpression of the encoded gene product.

The teachings of Genes V, Alitalo, and Merlino are confirmed in several more recent reports that also document the correlation between gene amplification and levels of protein. Appellants have submitted two more recent studies providing evidence that the teachings referred to above are still widely accepted by those of skill in the art. Orntoft *et al.* (Molecular and Cellular Proteomics, (2002) 1:37-45), studied transcript levels of 5600 genes in malignant bladder cancers which were linked to a gain/loss of chromosomal material using an array-based method. Orntoft *et al.* showed that there was a gene dosage effect and teach that "in general (18

of 23 cases) chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts." *Orntoft* at 37, column 1, abstract. In addition, Hyman *et al.* (Cancer Research, (2002) 62:6240-6245) used CGH analysis and cDNA microarrays to compare DNA copy numbers and mRNA expression of over 12,000 genes in breast cancer tumors and cell lines. They showed that there is "evidence of a prominent global influence of copy number changes on gene expression levels." *Hyman* at 6244, column 1, last paragraph.

Additional supportive teachings are also provided by Pollack *et al.* (PNAS, (2002) 99:12963-12968) who studied a series of primary human breast tumors and found that "[b]y analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer cell lines and tumors." *Pollack* at 12967 at column 1, (emphasis added). Their study found that "62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels." *Pollack* at 12963, column 1, abstract.

Appellants have also submitted two references which show that gene amplification leads to overexpression of the encoded protein. Bahnassy *et al.* (BMC Gastroenterology, (2004) 4:22-34), studied the amplification of *cyclin D1* and assessed the levels of the encoded protein by immunohistochemistry. Bahnassy *et al.* found a "significant correlation between *cyclin D1* gene amplification and protein overexpression." *Bahnassy* at 27, column 1. Similarly, Blancato *et al.* (British Journal of Cancer, (2004) 90(8), 1612-1619), report that overexpression of *c-myc* mRNA and c-Myc protein is related to the copy number of the *c-myc* amplification. *Blancato* at 1613, column 2. Bahnassy and Blancato demonstrate continued evidentiary support for the widely-accepted principle that gene amplification correlates with overexpression of the encoded protein. Presumably, amplification leads to overexpression of the encoded protein by increasing the amount of corresponding mRNA.

Together, these excerpts and articles collectively teach that it is more likely than not that gene amplification increases mRNA expression. This evidence establishes that there is a reasonable correlation between gene amplification and gene overexpression, and one of skill in

the art would believe, to a reasonable probability, that gene amplification would lead to increased gene expression.

This conclusion is supported by the declaration of Victoria Smith, Ph.D., an expert in the field of Molecular Biology. Exhibit B of her Declaration reports the results of microarray analysis conducted on the cDNA of the gene encoding PRO539 (DNA47465). The results indicate that the PRO539 mRNA is significantly overexpressed in eight of the twenty-six lung tumor samples tested compared to the normal lung tissue controls. That is the equivalent of nearly one in every three samples (31%). In addition, four out of five squamous cell lung carcinomas (80%) are significantly overexpressed (shown in bold). In contrast, only one of the seven individual normal lung tissue samples shows significant overexpression of the PRO539 gene (14%). The tumors analyzed in the microarray are not the same tumors reported in Example 16, and therefore it is not possible to directly relate gene amplification to mRNA overexpression. However, these data do support the conclusion that PRO539 mRNA is overexpressed in lung tumors, particularly squamous cell lung carcinomas.

Dr. Smith states that "[i]t is well-established in the art that overexpression of the mRNA for a gene is likely to lead to overexpression of the corresponding protein." *Smith Declaration* at ¶ 6. She explains:

> Given the known correlation between overexpression of a gene and the corresponding overexpression of the encoded protein, it is very likely that a similar number of lung tumors will overexpress the PRO539 protein, while very few normal lung tissue samples likely will. Together with the data reported in Example 16 that the gene encoding PRO539 is amplified in some lung tumors, including squamous cell lung carcinoma, the results reported in Exhibit B indicate that the PRO539 gene and protein, as well as antibodies to the encoded protein, can be used to differentiate some cancerous lung tissue, particularly squamous cell carcinoma, from normal lung tissue. *Smith Declaration* at ¶ 7 (emphasis in original).

Because not all lung tumors show overexpression of PRO539, it cannot be used to exclude a sample being tested as non-cancerous. However, Dr. Smith states that the PRO539 gene, protein, and corresponding antibodies are useful as a diagnostic tool for lung cancer, particularly squamous cell carcinoma, since a very high percentage of squamous cell lung carcinomas overexpress the gene and most likely the encoded protein, while very few normal lung samples do.

As discussed above, the Examiner relies on a single apparently contrary example of one gene reported by Pennica *et al.* to support the assertion that there is "no relationship between gene amplification and mRNA levels, since mRNA levels have no necessary correlation with gene amplification." *Office Action* at 4 (emphasis added). As Appellants discussed above, this possible contrary example is balanced by the report in Pennica that the amplification of the *WISP-1* gene by about two-fold led to overexpression of the corresponding mRNA. Thus, Pennica is at best inconclusive with one example of a positive correlation and one possible contrary example. If the example of *WISP-2* relied on by the Examiner is an artifact as the authors suggest is possible, the Pennica reference actually supports Appellants' position.

As stated above, the standard for utility is not absolute or even statistical certainty, but rather whether considering the evidence as a whole, one of skill in the art would be more likely than not to believe the asserted utility. Even if Pennica supported the PTO's argument, which it does not, one contrary example is not sufficient to prove that a person of skill in the art would have a reasonable doubt that gene amplification is not correlated to gene expression. Given the evidence provided by the Appellants which establishes that there is a correlation between gene amplification and mRNA expression, one of skill in the art would believe, to a reasonable probability, that the reported amplification of the PRO539 gene would lead to an increase in the level of PRO539 mRNA.

c. Appellants have established that generally there is a correlation between changes in mRNA expression levels and changes in the expression level of the encoded protein

Appellants turn to the next portion of their argument in support of their asserted utility – that it is well-established in the art that in most cases a change in the level of mRNA for a particular protein leads to a corresponding change in the level of the encoded protein. Given Appellants' evidence of two-fold amplification and overexpression of the PRO539 gene in certain lung and colon tumors, it is more likely than not that the PRO539 polypeptide is likewise overexpressed, and therefore the claimed antibodies are useful as diagnostic tools, particularly for lung and colon tumor.

In support of the assertion that changes in mRNA are positively correlated to changes in protein levels, Appellants submitted a Declaration by J. Christopher Grimaldi, an expert in the

field of cancer biology. As stated in paragraph 5 of the declaration, "Those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed...the gene product or polypeptide will also be over-expressed.... This same principal applies to gene under-expression." *Grimaldi Declaration* at ¶ 5. Further, "increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression." *Id.*

Appellants also submitted the declaration of Paul Polakis, Ph.D. an expert in the field of cancer biology. As stated in paragraph 6 of his declaration:

> Based on my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above [showing a positive correlation between mRNA levels and encoded protein levels in the vast majority of cases studied in relation to the present invention] and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, *it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. Polakis Declaration* at ¶ 6 (emphasis added).

Dr. Polakis acknowledges that there are published cases where such a correlation does not exist, but states that it is his opinion, based on over 20 years of scientific research, that "such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein." *Polakis Declaration* at ¶ 6.

As mentioned above, the Smith Declaration also states that "[i]t is well-established in the art that overexpression of the mRNA for a gene is likely to lead to overexpression of the corresponding protein." *Smith Declaration* at ¶ 6. She explains that "[g]iven the known correlation between overexpression of a gene and the corresponding overexpression of the encoded protein" it is very likely that those lung tumors that overexpress the PRO539 mRNA will also overexpress the PRO539 protein, and this indicates "that the PRO539 gene and protein, as well as antibodies to the encoded protein, can be used to differentiate some cancerous lung tissue, particularly squamous cell carcinoma, from normal lung tissue." *Smith Declaration* at ¶ 7.

The Declarations of Grimaldi, Polakis and Smith are supported by the teachings in Molecular Biology of the Cell, a leading textbook in the field (Bruce Alberts, *et al.*, Molecular Biology of the Cell (3rd ed. 1994) (hereinafter "Cell 3rd") and (4th ed. 2002) (hereinafter "Cell

4th"")). Figure 9-2 of Cell 3rd shows the steps at which eukaryotic gene expression can be controlled. The first step depicted is transcriptional control. Cell 3rd provides that "[f]or most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized." *Cell 3rd* at 403 (emphasis added). In addition, the text states that "Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made." *Cell 3rd* at 453 (emphasis added). Thus, as established in Cell 3rd, the predominant mechanism for regulating the amount of protein produced is by regulating transcription.

In Cell 4th, Figure 6-3 on page 302 illustrates the basic principle that there is a correlation between increased gene expression and increased protein expression. The accompanying text states that "a cell can change (or regulate) the expression of each of its genes according to the needs of the moment – most obviously by controlling the production of its mRNA." *Cell 4th* at 302 (emphasis added). Similarly, Figure 6-90 on page 364 of Cell 4th illustrates the path from gene to protein. The accompanying text states that while potentially each step can be regulated by the cell, "the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes." *Cell 4th* at 364 (emphasis added). This point is repeated on page 379, where the authors state that of all the possible points for regulating protein expression, "[f]or most genes transcriptional controls are paramount." *Cell 4th* at 379 (emphasis added).

Further support for Appellants' position can be found in the textbook, Genes VI, (Benjamin Lewin, Genes VI (1997)) which states "having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear that the overwhelming majority of regulatory events occur at the initiation of transcription." *Genes VI* at 847-848 (emphasis added).

Additional support is also found in Zhigang *et al.*, World Journal of Surgical Oncology 2:13, 2004. Zhigang studied the expression of prostate stem cell antigen (PSCA) protein and mRNA to validate it as a potential molecular target for diagnosis and treatment of human prostate cancer. The data showed "a high degree of correlation between PSCA protein and mRNA expression" *Zhigang* at 4. Of the samples tested, 81 out of 87 showed a high degree of correlation between mRNA expression and protein expression. The authors conclude that "it is

demonstrated that PSCA protein and mRNA overexpressed in human prostate cancer, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA." *Id.* at 6. Even though the correlation between mRNA expression and protein expression occurred in 93% of the samples tested, not 100%, the authors state that "PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor." *Id.* at 7.

Further, as noted above, Meric *et al.*, Molecular Cancer Therapeutics, vol. 1, 971-979 (2002), states the following:

> The **fundamental principle** of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells...[M]ost efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable to either DNA amplification or to differences in transcription. *Meric et al.* at 971 (emphasis added).

Those of skill in the art would not be focusing on differences in gene expression between cancer cells and normal cells if there were no correlation between gene expression and protein expression.

Together, the declarations of Grimaldi, Polakis, and Smith, the accompanying references, and the excerpts and references discussed above all establish that the accepted understanding in the art is that there is a reasonable correlation between changes in gene expression and changes in the level of the encoded protein. In contrast to this substantial amount of evidence supporting Appellants' position, the Examiner has two references, Meric *et al.* and Gökman-Polar *et al.* However, as discussed above, Meric and Gökman-Polar actually support Appellants' position, not the Examiner's. It is clear that when considered as a whole, the preponderance of the evidence clearly weighs in favor of Appellants.

d. The Examiner has failed to consider all of the evidence of record as a whole

Even assuming that the Examiner has established a *prima facie* case that one of skill in the art would doubt the truth of the asserted utility, Appellants have presented sufficient rebuttal evidence. The rebuttal evidence "will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (emphasis in original, citations omitted).

The Examiner has not considered the evidence as a whole. Instead, the Examiner has summarily stated that the Grimaldi and Polakis Declarations are "fundamentally flawed because they fail to provide specific evidence regarding Pro-539." *Office Action* mailed 9/8/04 at 11. In response to the Declaration of Dr. Smith which has specific evidence regarding PRO539, the Examiner is silent, completely ignoring her declaration. In response to the extensive references cited by the Appellants, the Examiner has summarily dismissed them as "nonspecific arguments" that do not relate to PRO539. *Office Action* at 16. He argues that "[n]one of the references demonstrate that there is a 'reasonable probability' that the Pro-539 protein is overexpressed or that antibodies to the Pro-539 protein itself have any utility." *Id.* Instead of weighing the evidence as a whole as required, the Examiner summarily states that "[a]s noted in the rejection, there are other articles which demonstrate that there is no necessary relationship [between nucleic acid overexpression and protein overexpression] for every protein." *Id.* (emphasis added).

Appellants have established that the PRO539 gene is amplified at least two-fold in the majority of lung and colon tumors tested. Appellants have submitted substantial evidence in the form of additional data, expert declarations, reference articles and textbooks, to support their assertion that it is more likely than not that this gene amplification leads to overexpression of the PRO539 mRNA and polypeptide. This makes the claimed antibodies to PRO539 polypeptide useful for diagnosing cancer, particularly lung and colon tumors. The Examiner has provided almost no evidence to the contrary. Given the overwhelming amount of evidence in support of Appellants' position, and the near absence of any evidence in support of the Examiner's position, when considered as a whole the evidence leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true.

8. ***The Courts have held that the Utility Requirement was Satisfied in Similar Cases***

The seminal decision interpreting the utility requirement of 35 U.S.C. § 101 is *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. 689 (1966). At issue in *Brenner* was a claim to "a chemical process which yields an already known product whose utility – other than as a possible object of scientific inquiry – ha[d] not yet been evidenced." *Id.* at 529, 148 U.S.P.Q. at 693. The Patent Office rejected the claimed process for lack of utility because the product produced by the

claimed process had no known use. *See id.* at 521-22, 148 U.S.P.Q. at 690. On appeal, the Court of Customs and Patent Appeals reversed, holding "where a claimed process produces a known product it is not necessary to show utility for the product." *Id.* at 522, 148 U.S.P.Q. at 691.

In reviewing the lower court's decision, the Court made its oft quoted statement that "[t]he basic quid pro quo contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility. Unless and until a process is refined and developed to this point – where specific benefit exists in currently available form – there is insufficient justification for permitting an applicant to engross what may prove to be a broad field." *Id.* at 534-35, 148 U.S.P.Q. at 695.

The first opinion of the C.C.P.A. applying *Brenner* was *In re Kirk*, 376 F.2d 936, 153 U.S.P.Q. 48 (C.C.P.A. 1967). The invention claimed in *Kirk* was a set of steroid derivatives said to have valuable biological properties and to be of value "in the furtherance of steroidal research and in the application of steroidal materials to veterinary or medical practice." *Id.* at 938, 153 U.S.P.Q. at 50. In affirming the claim rejection based on a lack of utility, the court held that the "nebulous expressions 'biological activity' or 'biological properties'" did not adequately convey how to use the claimed compounds. *Id.* at 941, 153 U.S.P.Q. at 52. The court also rejected appellants' supporting affidavit, stating, "the sum and substance of the affidavit appears to be that one of ordinary skill in the art would know 'how to use' the compounds to find out in the first instance whether the compounds are – or are not – in fact useful or possess useful properties, and to ascertain what those properties are." *Id.* at 942, 153 U.S.P.Q. at 53.

The instant case is clearly distinguishable from *Kirk*. In *Kirk*, the asserted utility for the claimed compounds was "a new class of compounds often possessing high biological activity" and "intermediates in the preparation of compounds with valuable biological properties...". *Id.* at 1120, 1121. Here, Appellants have asserted a much more specific utility than "biological activity" or "biological properties." Appellants have provided evidence of amplification and overexpression of the PRO539 gene in certain cancers and have shown that this evidence is reasonably correlated to overexpression of the PRO539 polypeptide in those same cancers, namely, lung and colon cancer. As Example 12 of the Utility Guidelines make clear, when a protein is differentially expressed in cancer compared to normal tissue, the protein and antibodies have utility in diagnosing the cancer.

Since the *Brenner* and *Kirk* decisions, the courts have continued to clarify what is sufficient to satisfy the utility requirement. Three more recent decisions are of particular relevance to the instant application: *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. 881 (C.C.P.A. 1980), *Cross v. Iizuka*, 753 F.2d 1040, 224 U.S.P.Q. 739 (Fed. Cir. 1985), and *Fujikawa v. Wattanasin*, 93 F.3d 1559, 39 U.S.P.Q. 2d 1895 (Fed. Cir. 1996).

The earliest of these cases, *Nelson v. Bowler*, involved an interference between two applications related to derivatives of naturally occurring prostaglandins (PG). *Nelson*, 626 F.2d at 854-55. The issue was whether Nelson had shown at least one utility for the compounds at issue to establish an actual reduction to practice. *Id.* at 855. The Appellants relied on two tests to prove practical utility: an *in vivo* rat blood pressure (BP) test and an *in vitro* gerbil colon smooth muscle stimulation (GC-SMS) test. In the BP test, the blood pressure of anesthetized rats was recorded on a polygraph chart to determine whether an injected compound had any effect. Responses were categorized as either a depressor (lowering) effect or a pressor (elevating) effect. *Id.* In the GC-SMS test a section of colon was excised from a freshly-killed gerbil for suspension in a physiological solution, and a lever arm was connected to the colon in such a way that any contraction was recorded as a polygraph trace. *Id.* The Board held that Nelson had not shown adequate proof of practical utility, characterizing the tests as "rough screens, uncorrelated with actual utility." *Id.* at 856.

On appeal the C.C.P.A. reversed, holding that the Board "erred in not recognizing that tests evidencing pharmacological activity may manifest a practical utility even though they may not establish a specific therapeutic use." *Id.* The Court stated that "practical utility" was characterized as a use of the claimed discovery in a manner which provides some immediate benefit to the public, establishing the following rule:

> Knowledge of the pharmacological activity of any compound is obviously beneficial to the public. It is inherently faster and easier to combat illnesses and alleviate symptoms when the medical profession is armed with an arsenal of chemicals having known pharmacological activities. Since it is crucial to provide researchers with an incentive to disclose pharmacological activities in as many compounds as possible, we conclude that adequate proof of <u>any</u> such activity constitutes a showing of practical utility. *Id.* (emphasis added).

The Court rejected Bowler's argument that the BP and GC-SMS tests are inconclusive showings of pharmacological activity since confirmation by statistically significant means did not occur until after the critical date. The Court stated that "a rigorous correlation is <u>not</u>

necessary where the test for pharmacological activity is reasonably indicative of the desired response." *Id.* (emphasis added). The Court concluded that a "reasonable correlation" between the observed properties and the suggested use was sufficient to establish practical utility. *Id.* at 857.

The sufficiency of a "reasonable correlation" in establishing utility was affirmed by the Court of Appeals for the Federal Circuit in *Cross v. Iizuka*, 753 F.2d 1040, 224 U.S.P.Q. 739 (Fed. Cir. 1985). In *Cross*, the subject of the interference before the Court was imidazole derivative compounds which inhibit the synthesis of thromboxane synthetase, an enzyme which leads to the formation of thromboxane A_2. At the time the applications were filed, thromboxane A_2 was postulated to be involved in platelet aggregation, which was associated with several deleterious conditions. *Id.* at 1042.

The question before the Board and reviewed by the Court was whether Iizuka was entitled to the benefit of his Japanese priority application. *Id.* The Japanese application disclosed that the imidazole derivatives showed strong inhibitory action for thromboxane synthetase from human or bovine platelet microsomes, an *in vitro* utility. *Id.* at 1043. Relying in part on *Nelson,* the Board held that tests evidencing pharmacological activity may manifest a practical utility even though they may not establish a specific therapeutic use, and concluded that the *in vitro* tests were sufficient to establish a practical utility. *Id.*

On appeal, Cross argued that the basic *in vitro* tests conducted in cellular fractions did not establish a practical utility for the claimed compounds, and that more sophisticated *in vitro* or *in vivo* tests were necessary to establish a practical utility. *Id.* at 1050. The Court rejected this argument, noting that adequate proof of any pharmaceutical activity constitutes a showing of practical utility. *Id.* The Court accepted the argument that initial testing of compounds is widely done *in vitro*:

> [*I*]*n vitro* results...are generally predictive of *in vivo* test results, i.e., there is a reasonable correlation therebetween. Were this not so, the testing procedures of the pharmaceutical industry would not be as they are. Iizuka has not urged, and rightly so, that there is an invariable exact correlation between *in vitro* test results and *in vivo* test results. Rather, Iizuka's position is that successful *in vitro* testing for a particular pharmacological activity establishes a significant probability that *in vivo* testing for this particular pharmacological activity will be successful. *Id.* (emphasis added).

The Court also noted that in previous decisions, its predecessor court had accepted evidence of *in vivo* utility as sufficient to establish practical utility. The Court reasoned that:

> This *in vivo* testing is but an intermediate link in a screening chain which may eventually lead to the use of the drug as a therapeutic agent in humans. We perceive no insurmountable difficulty, under appropriate circumstances, in finding that the first link in the screening chain, *in vitro* testing, may establish a practical utility for the compound in question. Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public, analogous to the benefit provided by the showing of an *in vivo* utility. *Id.* at 1051, *citing Nelson,* 626 F.2d at 856 (emphasis added).

Based on this reasoning, the Court affirmed the decision of the Board, stating that "based upon the relevant evidence as a whole, there is a reasonable correlation between the disclosed *in vitro* utility and an *in vivo* activity, and therefore a rigorous correlation is not necessary where the disclosure of pharmacological activity is reasonable based upon the probative evidence." *Id.* at 1050 (emphasis added). The Court therefore held that the disclosed *in vitro* utility was "sufficient to comply with the practical utility requirement of § 101." *Id.* at 1051.

The holdings of *Nelson* and *Cross* were more recently affirmed in *Fujikawa v. Wattanasin*, 93 F.3d 1559, 39 U.S.P.Q.2d 1895 (Fed. Cir. 1996). In *Fujikawa*, the Court again affirmed the notion that initial screens of compounds provide a practical utility even though they may not provide a therapeutic use because "'[i]t is inherently faster and easier to combat illnesses and alleviate symptoms when the medical profession is armed with an arsenal of chemicals having known pharmacological activities.'" *Id.* at 1564, *quoting Nelson,* 626 F.2d at 856. The Court noted that it may be difficult to predict whether novel compounds will exhibit pharmacological activity, and consequently testing is often required to establish practical utility. *Id.* However the Court went on to state:

> But the test results need not absolutely prove that the compound is pharmacologically active. All that is required is that the tests be "*reasonably* indicative of the desired [pharmacological] response." In other words, there must be a sufficient correlation between the tests and an asserted pharmacological activity so as to convince those skilled in the art, to a reasonable probability, that the novel compound will exhibit the asserted pharmacological behavior." *Id.* (internal citations omitted, underline emphasis added, italics in original).

On appeal, Fujikawa argued that Wattanasin had failed to establish an adequate correlation between the *in vitro* and *in vivo* results to permit Wattanasin to rely on positive *in*

vitro results to establish a practical utility. The Court stated that the Board relied on testimony from those skilled in the art that the *in vitro* results convinced the experts that the claimed compounds would exhibit the desired pharmacological activity when administered *in vivo*, including testimony that *in vivo* activity is typically highly correlatable to a compound's *in vitro* activity in the field. *Id.* at 1565. To overcome this evidence and counter the Board's decision, Fujikawa pointed to the testimony of its expert that "there is a reasonable element of doubt that some elements may be encountered which are active in the *in vitro* assay, but yet inactive in the *in vivo* assay." *Id.*

The Court rejected this argument: "Of course, it is possible that some compounds active *in vitro* may not be active *in vivo.* But, as our predecessor court in *Nelson* explained, a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices." *Id.* (emphasis added). The Court also rejected Fujikawa's reliance on two articles. The Court noted that while one article taught that "*in vitro* testing is sometimes not a good indicator of how potent a compound will be *in vivo,* it does imply that compounds which are active *in vitro* will normally exhibit some *in vivo* activity." *Id.* at 1566. Similarly, the Court noted that the second article expressly stated that "[f]or most substances, although not for all, the relative potency determined in *in vitro* ... parallels the *in vivo* activity." *Id.*

The Court concluded that the facts in the case were analogous to the ones in *Cross* where the court relied on a known reasonable correlation between *in vitro* tests and *in vivo* activity, and therefore affirmed the Board's decision that Wattanasin had established a practical utility with the *in vitro* results. *Id.* at 1565-66.

The *Nelson*, *Cross*, and *Fujikawa* cases are very similar to the present case. The reasoning of the courts in all three cases that "'[i]t is inherently faster and easier to combat illnesses and alleviate symptoms when the medical profession is armed with an arsenal of chemicals having known pharmacological activities'" applies to the asserted utility for the claimed antibodies. *Fujikawa*, 93 F.3d at 1564, *quoting Nelson,* 626 F.2d at 856; *see also Cross*, 753 F.2d at 1051 ("Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public, analogous to the benefit provided by the showing of an *in vivo* utility."). Like pharmaceutical compounds, nucleic acids, polypeptides, and antibodies which are associated with cancer will make it inherently faster and easier to combat cancer. The greater the

number of biological markers of cancer medical professionals have access to, the more accurate and detailed a diagnosis they can make. The determination that a gene is amplified and differentially expressed in cancer compared to normal tissue constitutes at least as significant a development in the field of cancer diagnostics as *in vitro* screening for pharmaceutical activity. *See Cross*, 753 F.2d at 1051 ("the first link in the screening chain, *in vitro* testing, may establish a practical utility for the compound in question. Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public.").

In addition, like *in vitro* tests in the pharmaceutical industry, those of skill in the field of biotechnology rely on the reasonable correlation that exists between gene amplification and mRNA and protein expression (see discussion supra). Were there no reasonable correlation between them, the techniques that measure gene levels such as microarray analysis, differential display, and quantitative PCR would not be so widely used by those in the art. *See Grimaldi Declaration* at ¶5. As in *Cross*, Appellants here do not argue that there is "an invariable exact correlation" between gene amplification, mRNA, and protein expression. *See Cross*, 753 F.2d at 1050. Instead, Appellants' position detailed above is that a measured change in gene amplification and gene expression in cancer cells establishes a "significant probability" that the expression of the encoded polypeptide in cancer will also be changed based on "a reasonable correlation therebetween." *Id.*; *see also Fujikawa*, 93 F.3d at 1565 ("a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices"); *Nelson*, 626 F.2d at 857 (holding that "a rigorous correlation is not necessary" and that a "reasonable correlation" will suffice).

Also of importance is the Court's rejection of the notion that any *in vitro* testing must be statistically significant to support a practical utility. *Nelson*, 626 F.2d at 857. Likewise, qualitative characterizations of a test compound as either increasing or decreasing blood pressure was acceptable. *Id.* at 855 (stating that responses were categorized as either a depressor (lowering) effect or a pressor (elevating) effect). This is similar to the data in Example 16, where the gene amplification was not evaluated for statistical significance, although Dr. Goddard's unchallenged declaration makes clear that the data are reliable and significant.

There are additional similarities. In *Fujikawa*, the Board and Court rejected the argument that there was no utility because there was no exact correlation between the *in vitro* and *in vivo*

results in spite of testimony and references establishing a lack of a "necessary" correlation. *Fujikawa*, 93 F.3d at 1565-66. Like the two references rejected by the Board and Court in *Fujikawa*, the references cited by the Examiner may suggest that the correlation between gene amplification and increases in mRNA levels and protein levels is not exact. But like *Fujikawa*, portions of the references also support Appellants' assertion, and Appellants have submitted the declaration of two experts in the field which state that those in the field rely on the correlation between changes in mRNA and protein. *See Grimaldi Declaration* at ¶ 5; *Polakis Declaration* at ¶ 6. Thus, as was the case in *Fujikawa*, although there may be some evidence that the correlation relied on is not exact, the declarations and numerous references submitted by Appellants are more than enough evidence to establish that there is a "reasonable correlation" between gene amplification, changes in mRNA levels and changes in the level of the encoded protein.

In conclusion, Appellants have asserted that the claimed antibodies are useful for the diagnosis of cancer, particularly lung and colon cancer based on the data in Example 16. This utility is far beyond the nebulous expressions "biological activity" or "biological properties" rejected in *In re Kirk*, 376 F.2d 936, 153 U.S.P.Q. 48 (C.C.P.A. 1967). Like *Nelson*, *Cross*, and *Fujikawa*, Appellants have asserted a utility which relies on a reasonable correlation between the data disclosed in the application and the asserted utility. The fact that there may be limited evidence that the correlation is not exact does not invalidate Appellants' showing of utility since the correlation need not be a rigorous or exact one. Considering the relevant evidence as a whole, Appellants have provided sufficient evidence to establish a reasonable correlation between gene amplification, changes in the level of mRNA and corresponding changes in the level of the encoded polypeptide. Therefore the claimed antibodies have a practical utility as diagnostic tools for lung and colon cancer.

9. *Utility – Conclusion*

Appellants' asserted utility for the claimed antibodies as diagnostic tools for cancer corresponds in scope to the subject matter sought to be patented and therefore "must be taken as sufficient to satisfy the utility requirement of § 101 for the entire claimed subject." *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. 288, 297 (C.C.P.A. 1974). The Examiner's unsupported arguments and references are not sufficient evidence to make a *prima facie* showing that "one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560,

1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). And even if the Examiner has established a *prima facie* case, Appellants have offered sufficient rebuttal evidence in the form of expert declarations and references, which, when considered as a whole, establish that it is more likely than not that the asserted utility is true. *See In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992) (stating that the evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard); *M.P.E.P.* § 2107.02, part VII ("evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true.") (emphasis in original).

Finally, the courts' decisions in similar cases make clear that the evidence provided by Appellants is sufficient to establish the asserted utility. The evidence does not need to be direct evidence, nor does it need to provide an exact correlation between the submitted evidence and the asserted utility. Instead, evidence which is "reasonably" correlated with the asserted utility is sufficient. *See Fujikawa*, 93 F.3d at 1565 ("a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices"); *Cross*, 753 F.2d at 1050 (same); *Nelson*, 626 F.2d at 857 (same). Considering the evidence as a whole in light of the relevant cases, the Board should find that Appellants have established at least one specific, substantial, and credible utility, and the Examiner's rejection of the pending claims as lacking utility should be reversed.

C. Enablement Rejection – Detailed Arguments

The second issue before the Board is whether Appellants have enabled the pending claims such that one of skill in the art would be able to make and use the claimed invention. The Examiner has rejected Claims 22-26 under 35 U.S.C. §112, first paragraph, arguing that the claimed subject matter was not described in the specification in such a way as to enable one skilled in the art to make and/or use the invention. *Office Action* at 7. The Examiner recites the factors for determining enablement from *In re Wands*, 858 F.2d 731, 8 U.S.P.Q. 2d 1400 (Fed. Cir. 1988), relying to a large extent on the arguments made above in support of the rejection for lack of utility.

Appellants submit that Claims 22-26 are enabled such that one of skill in the art could make and use the claimed antibodies without undue experimentation. It is well within those of

skill in the art to make antibodies which are specific to a disclosed sequence, and the Examiner has not stated otherwise. As described above, Appellants assert that the claimed antibodies are useful as diagnostic tools for cancer, particularly lung and colon cancer based amplification of the PRO539 gene and the resulting overexpression the PRO539 polypeptide in lung and colon tumors. Based on the disclosure in the application, one of skill in the art would be able to use the claimed antibodies as diagnostic tools to detect overexpression of the PRO539 polypeptide in suspected lung and colon tumors without undue experimentation.

1. *Enablement – Legal Standard*

An application enables the claims "if one skilled in the art, after reading the[] disclosure[], could practice the invention claimed ... without undue experimentation." *Chiron Corp. v. Genentech, Inc.*, 363 F.3d 1247, 1253 (Fed. Cir. 2004). "But the question of undue experimentation is a matter of degree. The fact that some experimentation is necessary does not preclude enablement; what is required is that the amount of experimentation 'must not be unduly extensive.'" *PPG Indus., Inc. v. Guardian Indus., Corp.*, 75 F.3d 1558, 1564 (Fed. Cir. 1996) (quoting *Atlas Powder Co. v. E.I. DuPont de Nemours & Co.*, 750 F.2d 1569, 1576 (Fed. Cir. 1984)).

While the application must enable one of ordinary skill in the art to practice the full scope of the claimed invention, "[t]hat is not to say that the specification itself must necessarily describe how to make and use every possible variant of the claimed invention, for the artisan's knowledge of the prior art and routine experimentation can often fill gaps, interpolate between embodiments, and perhaps even extrapolate beyond the disclosed embodiments, depending upon the predictability of the art." *AK Steel Corp. v. Sollac*, 344 F.3d 1234, 1244 (Fed. Cir. 2003).

2. *Enablement – Burden of Proof*

In order to make an enablement rejection, the PTO has the initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04. A specification teaching how to make and use the claimed subject matter must be taken as being in compliance with the enablement requirement unless there is a reason to doubt the objective truth of the statements contained therein which are relied on for enabling support. *Id.* It is incumbent for the PTO "to explain why it doubts the truth or accuracy of any statement in a

supporting disclosure and to back up assertions of its own with acceptable evidence or reasoning which is inconsistent with the contested statement." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971). This can be done "by making specific findings of fact, supported by the evidence, and then drawing conclusions based on these findings of fact." *Id.*

3. ***Enablement – Standard of Proof***

Once the examiner has weighed all the evidence and established a reasonable basis to question the enablement provided for the claimed invention, the burden falls on the applicant to present persuasive arguments, supported by suitable proofs where necessary, that one skilled in the art would be able to make and use the claimed invention using the application as a guide. *See M.P.E.P.* § 2164.05. "The evidence provided by applicant **need not be conclusive but merely convincing** to one skilled in the art." *Id.* (bold emphasis added, underline in original). "A declaration or affidavit is, itself, evidence that must be considered." *Id.* (emphasis in original).

The examiner must then "weigh all the evidence before him or her, including the specification and any new evidence supplied by applicant with evidence and/or sound scientific reasoning previously presented in the rejection and decide whether the claimed invention is enabled." *Id.* "The examiner should **never** make the determination based on personal opinion. The determination should always be based on the weight of all the evidence." *Id.* (emphasis in original).

4. ***Appellants' Specification Teaches How to Make and Use the Claimed Subject Matter***

The claimed subject matter relates to isolated antibodies which specifically bind to the polypeptide of SEQ ID NO: 7, PRO539. The specification discloses how to make the claimed antibodies, including Example 27 on page 127 of the specification which specifically describes the preparation of antibodies that bind PRO polypeptides. *Specification* at 90, line 20 through 97, line 4, and 127, line 13, through 128, line 1. In addition, the specification discloses that the claimed antibodies can be used in diagnostic assays to detect the expression of PRO539 in specific types of tissue. *Specification* at 98, lines 5-29. In light of the amplification of the PRO539 gene in a majority of lung and colon tumors tested, one of skill in the art would expect

the PRO539 polypeptide to overexpressed in these tumors as well. Therefore, given the teaching in the specification on how to make and use the claimed antibodies to detect expression of PRO539 in specific tissues, one of skill in the art would be enabled to practice the claimed invention without undue experimentation.

Because Appellants' specification teaches how to make and use the claimed subject matter, it must be taken as being in compliance with the enablement requirement unless there is a reason to doubt the objective truth of the statements contained therein which are relied on for enabling support. *See M.P.E.P.* § 2164.04. *Id.*

5. *The Examiner's Arguments and Appellants' Responses*

The PTO has the initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04. It is incumbent for the PTO "to explain why it doubts the truth or accuracy of any statement in a supporting disclosure and to back up assertions of its own with acceptable evidence or reasoning which is inconsistent with the contested statement." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971). This can be done "by making specific findings of fact, supported by the evidence, and then drawing conclusions based on these findings of fact." *Id.*

The Examiner has attempted to rely on the factors to be considered in determining whether a disclosure meets the enablement requirement as described in *In re Wands*, 858 F.2d 731, 8 U.S.P.Q. 2d 1400 (Fed. Cir. 1988). Appellants state the Examiner's arguments regarding each factor and respond to each in turn below.

a. The nature of the invention

The Examiner argues that the invention is "in a class of invention which the CAFC has characterized as 'the unpredictable arts such as chemistry and biology.' Mycogen Plant Sci., Inc. v. Monsanto Co., 243 F.3d 1316,1330 (Fed. Cir. 2001)." *Office Action* at 7.

While Appellants acknowledge that some aspects of biology can be unpredictable, the creation of specific antibodies to a disclosed antigen is not unpredictable. This is the very issue that was addressed in *In re Wands*, 858 F.2d 731, 8 U.S.P.Q. 2d 1400 (Fed. Cir. 1988). There the CAFC held that the disclosure was sufficient to enable one of skill in the art to make monoclonal antibodies to a disclosed antigen without undue experimentation. *Id.* at 740. If the

disclosure was sufficient at the time of filing of the *Wands* application in 1980, it cannot be that the art of making antibodies has become less predictable in the ensuing 25 years. Similarly, the use of antibodies to detect a target peptide such as PRO539 in a particular tissue is similarly well-established in the art. *See Specification* at 98, lines 5-18 and cited references which are incorporated by reference into the specification.

Thus, contrary to the Examiner's assertions, the nature of the invention weighs in favor of enablement as it is a predictable and well-established aspect of biology.

b. The breadth of the claims

The Examiner argues that "[t]he claims broadly encompass not only a particular PRO-539 antibody but also include any antibody which binds the polypeptide of SEQ ID NO: 7." *Office Action* at 7.

The broadest pending claim is Claim 22, which recites "An isolated antibody that *specifically* binds to the polypeptide of SEQ ID NO:7." Thus, only those antibodies which specifically bind the PRO539 polypeptide are encompassed by the claims. This is not a broad genus of antibodies, but is instead limited to a well-defined genus which binds a specific disclosed polypeptide.

Thus, contrary to the Examiner's assertions, the breath of the claims weighs in favor of enablement.

c. The quantity of experimentation

The Examiner argues that the quantity of experimentation would be extremely large "since there is significant variability in the activity of polypeptides and nucleic acids. It would require significant study to identify the actual function of the PRO-539 protein and nucleic acid, and identifying a use for this protein and resultant antibody would be an inventive, unpredictable and difficult undertaking in itself." *Office Action* at 8 (emphasis added).

As Appellants discussed above in response to similar arguments by the Examiner, the asserted use of the claimed antibodies does not rely on knowing the function of the PRO539 protein or nucleic acid. Instead, the claimed antibodies can be used as diagnostic tools for cancer by detecting the overexpression of the PRO539 polypeptide in lung and colon tumors. This use

in no way depends on the function of the PRO539 protein, and therefore experiments to determine the function of PRO539 are unnecessary.

In addition, as discussed above with respect to the nature of the invention, making antibodies which specifically bind to PRO539 does not require a great deal of experimentation, as the techniques are well-established in the art. *See In re Wands*, 858 F.2d 731 (reversing the Board's decision of non-enablement and holding that as of 1980, undue experimentation was not required to make high-affinity monoclonal antibodies to a target peptide).

Thus, contrary to the Examiner's assertions, the quantity of experimentation is minimal, and this factor weighs in favor of enablement.

d. The unpredictability of the art and the state of the prior art

Echoing the arguments regarding quantity of experimentation, the Examiner argues that "[t]he art is extremely unpredictable with regard to protein function in the absence of reliable information regarding the protein activity." *Office Action* at 8. The Examiner again cites Rost *et al.*, arguing that proteins with different sequences may have very different functions. He argues that because the function of the PRO539 protein is not known, "it is entirely unpredictable what function and activity will be found for this protein," and there is no prior art to resolve this ambiguity. *Id.*

The Examiner then repeats his arguments in support of the utility rejection, stating that "the overexpression of the nucleic acid is not relevant to the utility of the protein" and that "[t]here is no evidence that the protein itself is overexpressed." *Office Action* at 8. The Examiner repeats verbatim the arguments made above regarding a lack of a "necessary" connection between gene amplification and mRNA and protein levels, citing the Meric *et al.*, Gökman-Polar *et al.* and Pennica *et al.* references. *Office Action* at 8-10. The Examiner concludes by stating that "it is entirely unpredictable how one would use this antibody in any context whatsoever." *Office Action* at 10.

Appellants have respond at length to these arguments above, and refer the Board to the discussion of the utility rejection *supra*. To briefly summarize those arguments, Appellants have demonstrated that the references relied on by the Examiner do not support his position, and instead support the Appellants' arguments. Appellants have provided substantial evidence in the form of additional data, expert declarations, and references which show that it is well-established

and predictable that the more than two-fold amplification of the PRO539 gene in lung and colon tumors will result in overexpression of the PRO539 protein in those tumors. Therefore given the disclosure in the specification, one of skill in the art would be able to use the claimed antibodies as diagnostic tools for lung and colon cancer.

Thus, contrary to the Examiner's assertions, the prior art regarding the correlation between gene amplification and protein overexpression is sufficiently predictable that this factor weighs in favor of enablement.

e. Working examples

The Examiner asserts that the specification has no working examples that relate to the antibody or protein, and repeats his arguments that the data showing overexpression of the nucleic acid in certain cancer cell lines are not relevant. He argues that "there is no statistical showing that the overexpression of the nucleic acids is even significant" and "there is no showing that the results from nucleic acids have any correlation with the protein or antibody and the art cited above demonstrates that there is no presumption of such a correlation." *Office Action* at 10.

First, Appellants have submitted the declaration of Dr. Goddard which states that the data in Example 16 are sufficiently reliable and significant such that the PRO539 gene can be used as a cancer diagnostic tool. *See Goddard Declaration* at ¶7. Appellants remind the Board that "The evidence provided by applicant need not be conclusive but merely convincing to one skilled in the art" and that "[a] declaration or affidavit is, itself, evidence that must be considered." *M.P.E.P.* § 2164.05 (emphasis in original). The Examiner has not presented any evidence or argument in response to the Goddard Declaration, and therefore Appellants have established that the data regarding amplification of the PRO539 gene is significant. As to the Examiner's argument that there is no correlation between the gene amplification data and the protein and antibody based on the three cited references, this argument has been addressed above.

Appellants also note that contrary to the Examiner's assertion, Example 27 on page 127 of the specification describes the preparation of monoclonal antibodies which can specifically bind PRO polypeptides. *Specification* at 127, line 13 through 128, line 1. In addition, the specification teaches how to use antibodies to a PRO polypeptide to detect its expression in specific cell or tissue types. *Specification* at 98, lines 5-18.

Thus, contrary to the Examiner's assertions, there are sufficient working examples related to the claimed antibodies that this factor weighs in favor of enablement.

f. Guidance in the specification

The Examiner summarily states that "[t]he specification provides no specific or substantial uses for the PRO-539 antibody or protein." *Office Action* at 10.

Contrary to this assertion, the specification teaches that :

> [A]nti-PRO antibodies may be used in diagnostic assays for PRO, *e.g.*, detecting its expression in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases [Zola, Monoclonal Antibodies: A Manual of Techniques, CRC Press, Inc. (1987) pp. 147-158]. *Specification* at 98, lines 5-10.

In addition, Example 16 which shows a more than two-fold amplification of PRO539 in lung and colon tumors teaches that:

> Amplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as colon, lung, breast and other cancers and diagnostic determination of the presence of those cancers. *Specification* at 112, lines 0-3 (emphasis added).

This claim is supported by substantial evidence discussed above with respect to utility. Together, with the level of knowledge in the art and the disclosure of how to make antibodies which are specific to PRO539, the specification clearly teaches that the claimed antibodies to PRO539 are useful as diagnostic tools for lung and colon cancer, and how such use can be accomplished.

Thus, contrary to the Examiner's assertions, the specification provides sufficient guidance on how to use the claimed antibodies such that this factor weighs in favor of enablement.

g. Level of skill in the art

Appellants agree with the Examiner that the level of skill in the art is high, with the typical skilled artisan having a doctorate degree and several years of post-graduate work. Thus, this factor weighs heavily in favor of enablement.

6. ***Enablement – Conclusion***

For the reasons discussed above, all of the *In re Wand* factors weigh in favor of a finding of enablement. The Examiner has made no attempt to argue that one of skill in the art would be unable to make the claimed antibodies. The remainder of his arguments are essentially those made in support of the utility rejection which fail for the reasons articulated above. Therefore, the Examiner has failed to meet his initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04. The Examiner has failed to back up his assertions "with acceptable evidence or reasoning which is inconsistent with the contested statement [of enablement]." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971).

And even if the Examiner has met his burden, Appellants have presented persuasive arguments, supported by the evidence discussed above with respect to utility, that one skilled in the art would be able to make and use the claimed invention using the application as a guide. Appellants remind the Board that "[t]he evidence provided by applicant need not be conclusive but merely convincing to one skilled in the art." *M.P.E.P.* § 2164.05 (emphasis in original).

Considering all of the evidence provided by the Appellants to establish their asserted utility, along with the disclosure in the specification, the Board should find that Appellants have established that one of skill in the art would be able to make and use the claimed invention without undue experimentation, and the Examiner's rejection of the pending claims as lacking an enabling disclosure should be reversed.

D. Conclusion

In view of the arguments presented above, Appellants submit that the specification as filed provides a specific, substantial and credible utility for the claimed antibodies, and that one of skill in the art would be able to make and use the claimed antibodies without undue experimentation. Appellants therefore request that the Board reverse the Examiners rejections under 35 U.S.C. §§101 and 112.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: Sept. 21, 2005

By: _____
AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

1912684
090205

VIII. APPENDIX A – CLAIMS ON APPEAL

22. An isolated antibody that specifically binds to the polypeptide of SEQ ID NO:7.

23. The antibody of Claim 22 which is a monoclonal antibody.

24. The antibody of Claim 22 which is a humanized antibody.

25. The antibody of Claim 22 which is an antibody fragment.

26. The antibody of Claim 22 which is labeled.

IX. APPENDIX B – EVIDENCE

Attached hereto is a copy of the evidence cited in Appellants' Brief. The list of evidence below is accompanied by a statement setting forth where in the record that evidence was entered into the record by the Examiner.

Tab	Reference	Submitted	Entered
1	Declaration of Audrey Goddard, Ph.D.	Originally submitted with Amendment and Response to Office Action mailed July 15, 2004 at page 14	Entered by Examiner in Final Office Action mailed September 8, 2004
2	Benjamin Lewin, Genes V, 5th ed. 1994, pages 1196-1201	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 1	Entered by Examiner in Final Office Action mailed April 25, 2005
3	Alitalo (Med. Biol., (1984) 62:304-317)	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 2	Entered by Examiner in Final Office Action mailed April 25, 2005
4	Merlino et al. (J. Clin. Invest., (1985) 75:1077-1079)	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 3	Entered by Examiner in Final Office Action mailed April 25, 2005
5	Orntoft et al. (Molecular and Cellular Proteomics, (2002) 1:37-45),	Originally submitted with Amendment and Response to Final Office Action mailed December 6, 2004 as Exhibit 1	Entered by Examiner in Advisory Action mailed December 20, 2004
6	Hyman et al. (Cancer Research, (2002) 62:6240-6245)	Originally submitted with Amendment and Response to Final Office mailed December 6, 2004 Action as Exhibit 2	Entered by Examiner in Advisory Action mailed December 20, 2004
7	Pollack et al. (PNAS, (2002) 99:12963-12968)	Originally submitted with Amendment and Response to Final Office Action mailed December 6, 2004 as Exhibit 3	Entered by Examiner in Advisory Action mailed December 20, 2004
8	Bahnassy et al. (BMC Gastroenterology, (2004) 4:22-34),	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 7	Entered by Examiner in Final Office Action mailed April 25, 2005

9	Blancato et al. (British Journal of Cancer, (2004) 90(8), 1612-1619)	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 8	Entered by Examiner in Final Office Action mailed April 25, 2005
10	Declaration of Victoria Smith, Ph.D.	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 13	Entered by Examiner in Final Office Action mailed April 25, 2005
11	Declaration by J. Christopher Grimaldi,	Originally submitted with Amendment and Response to Office Action mailed July 15, 2004 at page 15	Entered by Examiner in Final Office Action mailed September 8, 2004
12	Declaration of Paul Polakis, Ph.D.	Originally submitted with Amendment and Response to Office Action mailed July 15, 2004 at page 15	Entered by Examiner in Final Office Action mailed September 8, 2004
13	Bruce Alberts, *et al.*, Molecular Biology of the Cell (3rd ed. 1994)	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 9	Entered by Examiner in Final Office Action mailed April 25, 2005
14	Bruce Alberts, *et al.*, Molecular Biology of the Cell (4th ed. 2002)	Originally submitted with Amendment and Response to Final Office Action mailed December 6, 2004 as Exhibit 4	Entered by Examiner in Advisory Action mailed December 20, 2004
15	Benjamin Lewin, Genes VI (1997), pages 847-848	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 11	Entered by Examiner in Final Office Action mailed April 25, 2005
16	Zhigang et al., World Journal of Surgical Oncology 2:13, 2004	Originally submitted with Submission Filed With RCE mailed March 7, 2005 as Exhibit 12	Entered by Examiner in Final Office Action mailed April 25, 2005
17	Meric et al., Molecular Cancer Therapeutics, vol. 1, 971-979 (2002)		Cited by Examiner in Office Action mailed September 8, 2004

X. APPENDIX C – RELATED PROCEEDINGS

There are no decisions rendered by a court or the Board in any related proceedings identified above.

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ashkenazi et al. Serial No.: 09/903,925 Filed: July 11, 2001 For: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS	Group Art Unit: 1647 Examiner: Fozia Hamid CERTIFICATE OF MAILING I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231, on Date

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.

2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

Serial No.: *
Filed: *

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl., 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

Serial No.: *
Filed: *

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003
Date

Audrey Goddard
Audrey D. Goddard, Ph.D.

AUDREY D. GODDARD, Ph.D.

Genentech, Inc.
1 DNA Way
South San Francisco, CA, 94080
650.225.6429
goddarda@gene.com

110 Congo St.
San Francisco, CA, 94131
415.841.9154
415.819.2247 (mobile)
agoddard@pacbell.net

PROFESSIONAL EXPERIENCE

Genentech, Inc. **1993-present**
South San Francisco, CA

2001 - present **Senior Clinical Scientist**
Experimental Medicine / BioOncology, Medical Affairs

Responsibilities:
- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

Interests:
- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

1998 - 2001 **Senior Scientist**
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities:
- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a $750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

Research:
- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

1993 - 1998 Scientist
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities
- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

Research:
- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with Idiopathic short stature*

Imperial Cancer Research Fund **1989-1992**
London, UK with Dr. Ellen Solomon

6/89 –12/92: Postdoctoral Fellow
- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

McMaster University **1983**
Hamilton, Ontario, Canada with Dr. G. D. Sweeney

5/83 – 8/83: NSERC Summer Student
- *In vitro* metabolism of β-naphthoflavone in C57Bl/6J and DBA mice

EDUCATION

Ph.D. "Phenotypic and genotypic effects of mutations in the human retinoblastoma gene." **Supervisor:** Dr. R. A. Phillips	University of Toronto Toronto, Ontario, Canada. Department of Medical Biophysics.	1989
Honours B.Sc "The *in vitro* metabolism of the cytochrome P-448 inducer β-naphthoflavone in C57BL/6J mice." **Supervisor:** Dr. G. D. Sweeney	McMaster University, Hamilton, Ontario, Canada. Department of Biochemistry	1983

ACADEMIC AWARDS

Imperial Cancer Research Fund Postdoctoral Fellowship	1989-1992
Medical Research Council Studentship	1983-1988
NSERC Undergraduate Summer Research Award	1983
Society of Chemical Industry Merit Award (Hons. Biochem.)	1983
Dr. Harry Lyman Hooker Scholarship	1981-1983
J.L.W. Gill Scholarship	1981-1982
Business and Professional Women's Club Scholarship	1980-1981
Wyerhauser Foundation Scholarship	1979-1980

INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function, Litchfield Park, AZ, USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76th Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

Goddard A, Godowski PJ and Gurney AL. NL3 Tie ligand homologue nucleic acids. Patent Number: 6,426,218. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,4137,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date of Patent: May 14, 2002.

Goddard A, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homologue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard** A, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 19, 2002.

Goddard A, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,873. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesunheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesunheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

PUBLICATIONS

Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD**. Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.

Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausek SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.

Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.

Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.

Lee J. Ho WH. Maruoka M. Corpuz RT. Baldwin DT. Foster JS. **Goddard AD**. Yansura DG. Vandlen RL. Wood WI. Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.

Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.

Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8950-8954.

Guo S, Yamaguchi Y, Schilbach S, Wada T, Lee J, **Goddard A**, French D , Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.

Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.

Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.

Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.

Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused, a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.

Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* **274**(15): 10287-10292.

Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* **9**(4): 215-218.

Ridgway JBB, Ng E, Kern JA ,Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* 59: 2718-2723.

Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**(6712): 699-703.

Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* **95**(25): 14717-14722.

Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**(6699): 284-288.

Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* **16**(7): 677-681.

Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* **8**(9): 525-528.

Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature*. **391**(6662): 90-92.

Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology*. **7**(12): 1003-1006.

Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* **19**: 15–26.

Sheridan JP, Marsters SA, Pitti RM, Gurney A., Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* **277** (5327): 818-821.

Goddard AD, Dowd P, Chernausek S, Geffner M, Gertner J, Hintz R, Hopwood N, Kaplan S, Plotnick L, Rogol A, Rosenfield R, Saenger P, Mauras N, Hershkopf R, Angulo M and Attie, K. (1997) Partial growth hormone insensitivity: The role of growth hormone receptor mutations in idiopathic short stature. *J. Pediatr.* **131**: S51-55.

Klein RD, Sherman D, Ho WH, Stone D, Bennett GL, Moffat B, Vandlen R, Simmons L, Gu Q, Hongo JA, Devaux B, Poulsen K, Armanini M, Nozaki C, Asai N, **Goddard A**, Phillips H, Henderson CE, Takahashi M and Rosenthal A. (1997) A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. **387**(6634): 717-21.

Stone DM, Hynes M, Armanini M, Swanson TA, Gu Q, Johnson RL, Scott MP, Pennica D, **Goddard A**, Phillips H, Noll M, Hooper JE, de Sauvage F and Rosenthal A. (1996) The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**(6605): 129-34.

Marsters SA, Sheridan JP, Donahue CJ, Pitti RM, Gray CL, **Goddard AD**, Bauer KD and Ashkenazi A. (1996) Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF-kappa β. *Current Biology* **6**(12): 1669-76.

Rothe M, Xiong J, Shu HB, Williamson K, **Goddard A** and Goeddel DV. (1996) I-TRAF is a novel TRAF-interacting protein that regulates TRAF-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **93**: 8241-8246.

Yang M, Luoh SM, **Goddard A**, Reilly D, Henzel W and Bass S. (1996) The bglX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic beta-glucosidase. *Microbiology* **142**: 1659-65.

Goddard AD and Black DM. (1996) Familial Cancer in Molecular Endocrinology of Cancer. Waxman, J. Ed. Cambridge University Press, Cambridge UK, pp.187-215.

Treanor JJS, Goodman L, de Sauvage F, Stone DM, Poulson KT, Beck CD, Gray C, Armanini MP, Pollocks RA, Hefti F, Phillips HS, **Goddard A**, Moore MW, Buj-Bello A, Davis AM, Asai N, Takahashi M, Vandlen R, Henderson CE and Rosenthal A. (1996) Characterization of a receptor for GDNF. *Nature* **382**: 80-83.

Klein RD, Gu Q, **Goddard A** and Rosenthal A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**: 7108-7113.

Winslow JW, Moran P, Valverde J, Shih A, Yuan JQ, Wong SC, Tsai SP, **Goddard A**, Henzel WJ, Hefti F and Caras I. (1995) Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**: 973-981.

Bennett BD, Zeigler FC, Gu Q, Fendly B, **Goddard AD**, Gillett N and Matthews W. (1995) Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Htk. *Proc. Natl. Acad. Sci. USA* **92**: 1866-1870.

Huang X, Yuang J, **Goddard A**, Foulis A, James RF, Lernmark A, Pujol-Borrell R, Rabinovitch A, Somoza N and Stewart TA. (1995) Interferon expression in the pancreases of patients with type I diabetes. *Diabetes* **44**: 658-664.

Goddard AD, Yuan JQ, Fairbairn L, Dexter M, Borrow J, Kozak C and Solomon E. (1995) Cloning of the murine homolog of the leukemia-associated PML gene. *Mammalian Genome* **6**: 732-737.

Goddard AD, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.

Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominantly brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.

Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.

Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.

Goddard AD and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.

Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Fioretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.

Goddard AD, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S–119S.

Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.

Foulkes W, **Goddard A** and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.

Goddard AD, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.

Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.

Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the β–subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.

Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a *NotI* linking library from human chromosome 17q. *Genomics* **10**: 477–480.

Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* 46: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* 62: 394-408.

Goddard AD, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* 37: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* 340: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

Goddard AD, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* 8: 2082-2088.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1986) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci.* USA 83: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-*myc* expression. *Nature* 322: 555-557.

Goddard AD, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* 152: 31-38.

RESEARCH/

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger†, P. Sean Walsh and Robert Griffith

Roche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608; †Chiron Corporation, 1400 53rd St., Emeryville, CA 94608; *Corresponding author

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR[1] to clinical diagnostics are well known[2,3], it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases[4] made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization[5,6], gel electrophoresis with or without use of restriction digestion[7,8], HPLC[9], or capillary electrophoresis[10]. These methods are labor intense, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of "carryover" false positives in subsequent testing[11].

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.[12], developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.[13], developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to ds-DNA[14–16]. As outlined in Figure 1, a prototypic PCR

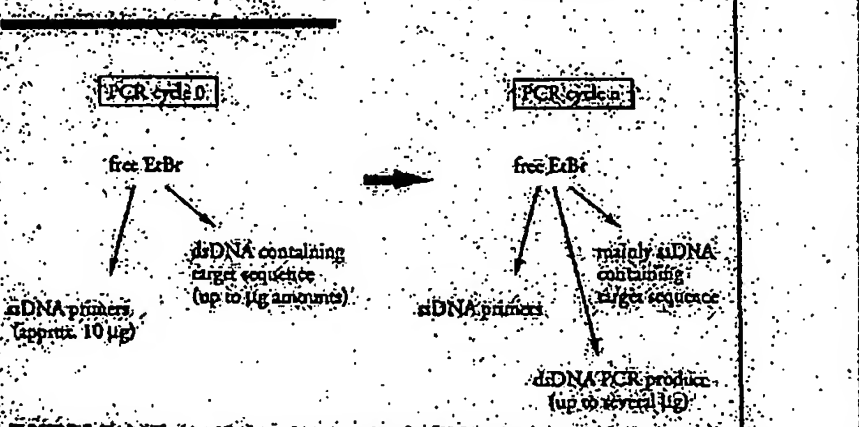


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

FIGURE 2 Gel electrophoresis of PCR amplification products of the human, nuclear gene, HLA-DQα, made in the presence of increasing amounts of EtBr (up to 8 μg/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

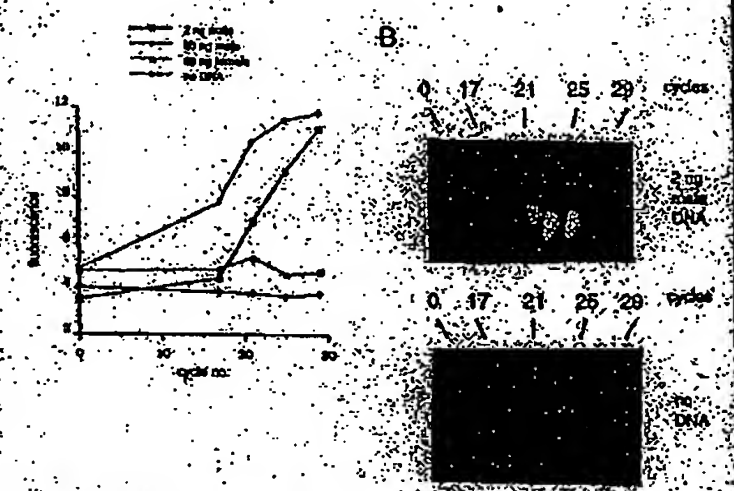


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μg/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA[17] to micrograms per PCR[16]. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the affect of EtBr in PCR, amplifications of the human HLA-DQα gene[15] were performed with the dye present at concentrations from 0.06 to 8.0 μg/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μg/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μg/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome[20]. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β-globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μg/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β-globin gene[21]. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3′ nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3′ nucleotide of the primer is complementary to the β-globin allele present[21,22].

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type β-globin individual (AA); from a heterozygous sickle β-globin individual (AS); and from a homozygous sickle β-globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β-globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β-globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β-globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells[6]. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

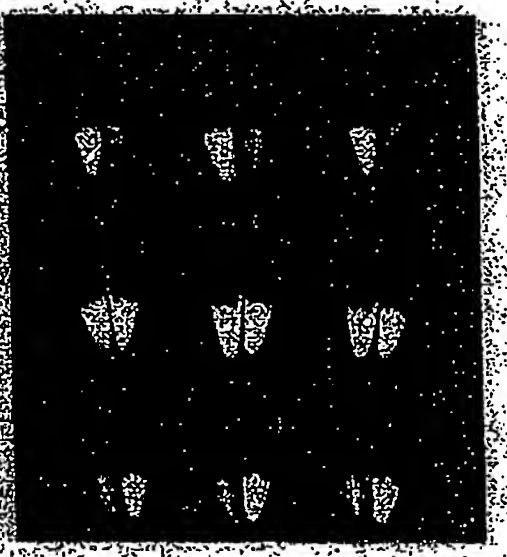


FIGURE 4 UV photography of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β-globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

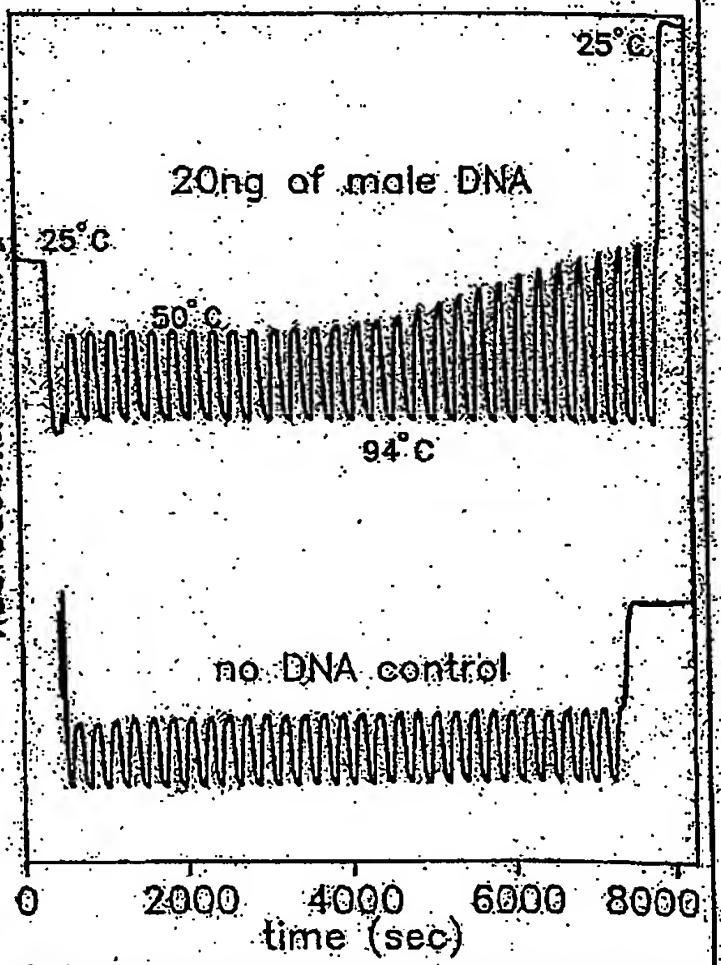


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested primer amplifications that take place in a single tube[7], and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins[20]. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer[24].

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format[25]. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader[26].

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQα gene amplifications containing EtBr. PCRs were set up in 100 µl volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM $MgCl_2$; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQα gene specific oligonucleotide primers GH26 and GH27[19] and approximately 10^7 copies of DQα PCR product diluted from a previous reaction. Ethidium bromide (EtBr, Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min. denaturation and 60°C for 30 sec. annealing and 72°C for 30 sec. extension.

Y-chromosome specific PCR. PCRs (100 µl total reaction volume) containing 0.5 µg/ml EtBr were prepared as described for HLA-DQα, except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2[20], and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min. and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β-globin gene PCR. Amplifications of 100 µl volume using 0.5 µg/ml of EtBr were prepared as described for HLA-DQα above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HβI4A (wild-type globin-specific primers) or HGP2/HβI4S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.[21]. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min. and 55°C for 1 min. using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.[21] to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griot, Inc., Irvine, CA) to exclude second order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluormetric measurements and Heatherbell Fong for editing this manuscript.

References

1. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. CSHSQB 51:263–273.
2. White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. Trends Genet. 5:185–189.
3. Erlich, H. A., Gelfand, D. and Sninsky, J. J. 1991. Recent advances in the polymerase chain reaction. Science 252:1643–1651.
4. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491.
5. Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. USA 86:6230–6234.
6. Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. J. Virol. 61:1690–1694.
7. Chehab, F. F., Doherty, M., Cai, S. R., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemias. Nature 329:293–294.
8. Horn, G. T., Richards, B. and Klinger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. Nuc. Acids Res. 16:2140.
9. Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. Biotechniques 8:546–555.
10. Heiger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. J. Chromatogr. 516:33–48.
11. Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. Nature 339:237–238.
12. Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. Proc. Natl. Acad. Sci. USA 86:9178–9182.
13. Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5′ to 3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Natl. Acad. Sci. USA 88:7276–7280.
14. Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. Anal. Biochem. 94:259–264.
15. Kapuscinski, J. and Szer, W. 1979. Interactions of 4′,6-diamidine-2-phenylindole with synthetic polynucleotides. Nuc. Acids Res. 6:3519–3534.
16. Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-rich DNA duplex studied in solution by ^{1}H NMR spectroscopy. Nuc. Acids Res. 18:3765–3762.
17. Li, H. H., Gyllensten, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414–417.
18. Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Sninsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. J. Infect. Dis. 158:1158.
19. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324:163–166.
20. Kogan, S. C., Doherty, M. and Gitschier, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N. Engl. J. Med. 317:985–990.
21. Wu, D. Y., Ugozzoli, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA 86:2757–2760.
22. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nuc. Acids Res. 18:999–1005.
23. Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Submitted.
24. Higuchi, R. 1989. Using PCR to engineer DNA. p. 61–70. *In* PCR Technology. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
25. Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Picotra, R., Williams, J. F. and Woudenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. Biotechniques 10:102–103, 106–112.
26. Tumosa, N. and Kahan, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. J. Immun. Meth. 116:59–63.

Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5′ nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5′ → 3′ nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3′-end nucleotide. In all cases, the reporter dye was attached to the 5′ end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3′-end nucleotide exhibited a larger signal in the 5′ nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3′-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridization probes.

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.[1] The assay exploits the 5′ → 3′ nucleolytic activity of *Taq* DNA polymerase[2,3] and is diagramed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5′ end, and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).[4,5] During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5′ → 3′ nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5′ nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.[6] Because of this, it has been assumed that the quencher dye must be attached near the 5′ end. Surprisingly, we have found that attaching a rhodamine dye at the 3′ end of a probe still provides adequate quenching for the probe to perform in the 5′ nuclease PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5′ end and a quencher dye on the 3′ end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

MATERIALS AND METHODS

Oligonucleotides

Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3′ blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5′ end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3′ end. Following deprotection and ethanol precipitation, TAMRA NHS ester was coupled to the LAN-containing oligonucleotide in 250

4:357–362©1995 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/95 $5.00

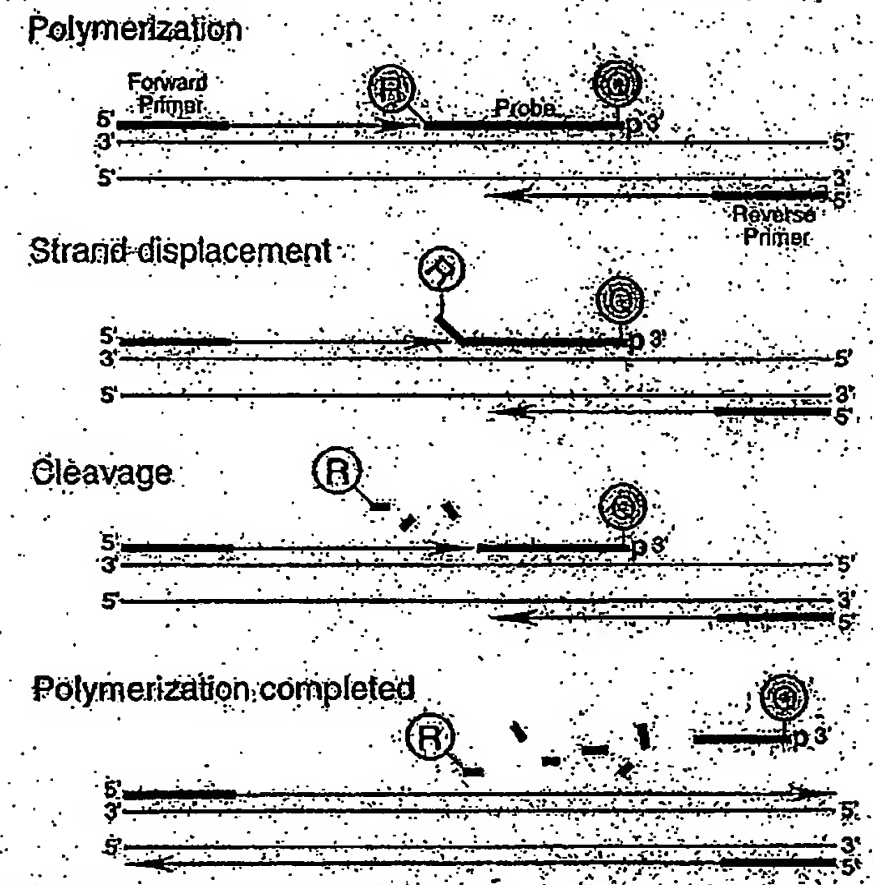


FIGURE 1 Diagram of 5′ nuclease assay. Stepwise representation of the 5′→3′ nucleolytic activity of *Taq* DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mM Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C_8 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the LAN-TAMRA moiety. For example, probe A1-7 has sequence A1 with LAN-TAMRA at nucleotide position 7 from the 5′ end.

PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris·HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil *N*-glycosylase (Perkin-Elmer), and 1.25 unit of AmpliTaq DNA polymerase (Perkin-Elmer). A 295-bp segment from exon 3 of the human β-actin gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.)[7] was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.[8] Actin amplification reactions contained 4 mM $MgCl_2$, 20 ng of human genomic DNA, 50 nM A1 or A3 probe, and 300 nM each primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 3.5 mM $MgCl_2$, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer F119, and 200 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer TaqMan LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 485-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

Name	Type	Sequence
F119	primer	ACCCACAGGAAGTGATCACCACTC
R119	primer	ATGTCGCGTTCCGGCTGACGTTCTGC
P2	probe	TCGCATTACTGATCGTTGCCAACCAGTp
P2C	complement	GTACTGGTTGGCAACGATCAGTAATGCGATG
P5	probe	CGGATTTGCTGGTATCTATGACAAGGATp
P5C	complement	TTCATCCTTGTCATAGATACCAGCAAATCCG
AFP	primer	TCACCCACACTGTGCCCATCTACGA
ARP	primer	CAGCGGAACCGCTCATTGCCAATGG
A1	probe	ATGCCCTCCCCCATGCCATCCTGCGTp
A1C	complement	AGACGCAGGATGGCATGGGGGAGGGCATAC
A3	probe	CGCCGTGGACTTGGAGCAAGAGATp
A3C	complement	CCATCTCTTGCTCGAAGTCCAGGGCGAC

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5′→3′ direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used in the 5′ nuclease assay, and complement used to hybridize to the corresponding probe. For the probes, the underlined base indicates a position where LAN with TAMRA attached was substituted for a T. (p) The presence of a 3′ phosphate on each probe.

A1-2	RAQGCCCTCCCCCATGCCATCCTCGGTp
A1-7	RATGCCCQCCCCCATGCCATCCTCGGTp
A1-14	RATGCCCTCCCCCAQGCCATCCTGCGTp
A1-19	RATGCCCTCCCCCATGCCAQCCTGCGTp
A1-22	RATGCCCTCCCCCATGCCATCCQCGGTp
A1-26	RATGCCCTCCCCCATGCCATCCTGCGQp

Probe	518 nm		582 nm		RQ⁻	RQ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A1-2	25.5 ± 2.1	32.7 ± 1.9	38.2 ± 3.0	38.2 ± 2.0	0.67 ± 0.01	0.86 ± 0.08	0.19 ± 0.06
A1-7	53.5 ± 4.3	395.1 ± 21.4	108.5 ± 6.3	110.3 ± 6.3	0.49 ± 0.03	3.58 ± 0.17	3.09 ± 0.18
A1-14	127.0 ± 4.9	403.5 ± 19.1	109.7 ± 6.3	93.1 ± 6.3	1.16 ± 0.02	4.34 ± 0.15	3.18 ± 0.15
A1-19	187.5 ± 17.9	422.7 ± 7.7	70.3 ± 7.4	73.0 ± 2.8	2.67 ± 0.05	5.80 ± 0.15	3.13 ± 0.16
A1-22	224.6 ± 9.4	482.2 ± 43.6	100.0 ± 4.0	96.2 ± 9.6	2.25 ± 0.03	5.02 ± 0.11	2.77 ± 0.12
A1-26	160.2 ± 8.9	454.1 ± 18.4	93.1 ± 5.4	90.7 ± 3.2	1.72 ± 0.02	5.01 ± 0.09	3.29 ± 0.08

FIGURE 2 Results of 5′ nuclease assay comparing β-actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average ± 1 s.d. for six reactions run without added template (no temp.) and six reactions run with template (+temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ⁻ and RQ⁺ values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, ΔRQ is calculated by subtracting the RQ value of the no-template control (RQ^-) from the RQ value for the complete reaction including template (RQ^+).

RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5′ nuclease PCR assay. These probes hybridize to a target sequence in the human β-actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β-actin gene containing the target sequence. Performance in the 5′ nuclease PCR assay is monitored by the magnitude of ΔRQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a ΔRQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5′ end, there is insufficient room for *Taq* polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable ΔRQ values that are clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ^- depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ^- values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3′ end to quench 6-FAM on the 5′ end, three additional pairs of probes were tested in the 5′ nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3′ end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3′ quencher exhibits a ΔRQ value that is considerably higher than for the probe with the internal quencher. The RQ^- values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3′ end of an oligonucleotide can quench efficiently the

TABLE 2 Results of 5′ Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3′-terminal Nucleotide

Probe	518 nm		582 nm		RQ⁻	RQ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A3-6	54.6 ± 3.2	84.8 ± 3.7	116.2 ± 6.4	115.6 ± 2.5	0.47 ± 0.02	0.73 ± 0.03	0.26 ± 0.04
A3-24	72.1 ± 2.9	236.5 ± 11.1	84.2 ± 4.0	90.2 ± 3.8	0.86 ± 0.02	2.62 ± 0.05	1.76 ± 0.05
P2-7	82.8 ± 4.4	384.0 ± 34.1	105.1 ± 6.4	120.4 ± 10.2	0.79 ± 0.02	3.19 ± 0.16	2.40 ± 0.16
P2-27	113.4 ± 6.6	555.4 ± 14.1	140.7 ± 8.5	118.7 ± 4.8	0.81 ± 0.01	4.68 ± 0.10	3.88 ± 0.10
P5-10	77.5 ± 6.5	244.4 ± 15.9	86.7 ± 4.3	95.8 ± 6.7	0.89 ± 0.05	2.55 ± 0.06	1.66 ± 0.08
P5-28	64.0 ± 5.2	333.6 ± 12.1	100.6 ± 6.1	94.7 ± 6.3	0.63 ± 0.02	3.53 ± 0.12	2.89 ± 0.13

Reactions containing the indicated probes and calculations were performed as described in Material and Methods and in the legend to fig. 2.

fluorescence of a reporter dye on the 5′ end. The degree of quenching is sufficient for this type of oligonucleotide to be used as a probe in the 5′ nuclease PCR assay.

To test the hypothesis that quenching by a 3′ TAMRA depends on the flexibility of the oligonucleotide, fluorescence was measured for probes in the single-stranded and double-stranded states. Table 3 reports the fluorescence observed at 518 and 582 nm. The relative degree of quenching is assessed by calculating the RQ ratio. For probes with TAMRA 6–10 nucleotides from the 5′ end, there is little difference in the RQ values when comparing single-stranded with double-stranded oligonucleotides. The results for probes with TAMRA at the 3′ end are much different. For these probes, hybridization to a complementary strand causes a dramatic increase in RQ. We propose that this loss of quenching is caused by the rigid structure of double-stranded DNA, which prevents the 5′ and 3′ ends from being in proximity.

When TAMRA is placed toward the 3′ end, there is a marked Mg^{2+} effect on quenching. Figure 3 shows a plot of observed RQ values for the A1 series of probes as a function of Mg^{2+} concentration. With TAMRA attached near the 5′ end (probe A1-2 or A1-7), the RQ value at 0 mM Mg^{2+} is only slightly higher than RQ at 10 mM Mg^{2+}. For probes A1-19, A1-22, and A1-26, the RQ values at 0 mM Mg^{2+} are very high, indicating a much reduced quenching efficiency. For each of these probes, there is a marked decrease in RQ at 1 mM Mg^{2+} followed by a gradual decline as the Mg^{2+} concentration increases to 10 mM. Probe A1-14 shows an intermediate RQ value at 0 mM Mg^{2+} with a gradual decline at higher Mg^{2+} concentrations. In a low-salt environment with no Mg^{2+} present, a single-stranded oligonucleotide would be expected to adopt an extended conformation because of electrostatic repulsion. The binding of Mg^{2+} ions acts to shield the negative charge of the phosphate backbone so that the oligonucleotide can adopt conformations where the 3′ end is close to the 5′ end. Therefore, the observed Mg^{2+} effects support the notion that quenching of a 5′ reporter dye by TAMRA at or near the 3′ end depends on the flexibility of the oligonucleotide.

DISCUSSION

The striking finding of this study is that it seems the rhodamine dye TAMRA, placed at any position in an oligonucleotide, can quench the fluorescent emission of a fluorescein (6-FAM) placed at the 5′ end. This implies that a single-stranded, double-labeled oligonucleotide must be able to adopt conformations where the TAMRA is close to the 5′ end. It should be noted that the decay of 6-FAM in the excited state requires a certain amount of time. Therefore, what matters for quenching is not the average distance between 6-FAM and TAMRA but, rather, how close TAMRA can get to 6-FAM during the lifetime of the 6-FAM excited state. As long as the decay time of the excited state is relatively long compared with the molecular motions of the oligonucleotide, quenching can occur. Thus, we propose that TAMRA at the 3′ end, or any other position, can quench 6-FAM at the 5′ end because TAMRA is in proximity to 6-FAM often enough to be able to accept energy transfer from an excited 6-FAM.

Details of the fluorescence measurements remain puzzling. For example, Table 3 shows that hybridization of probes A1-26, A3-24, and P5-28 to their complementary strands not only causes a large increase in 6-FAM fluorescence at 518 nm but also causes a modest increase in TAMRA fluorescence at 582 nm. If TAMRA is being excited by energy transfer from quenched 6-FAM, then loss of quenching attributable to hybridization should cause a decrease in the fluorescence emission of TAMRA. The fact that the fluorescence emission of TAMRA increases indicates that the situation is more complex. For example, we have anecdotal evidence that the bases of the oligonucleotide, especially G, quench the fluorescence of both 6-FAM and TAMRA to some degree. When double-stranded, base-pairing may reduce the ability of the bases to quench. The primary factor causing the quenching of 6-FAM in an intact probe is the TAMRA dye. Evidence for the importance of TAMRA is that 6-FAM fluorescence remains relatively unchanged when probes labeled only with 6-FAM are used in the 5′ nuclease PCR assay (data not shown). Secondary effectors of fluorescence, both before and after cleavage of the probe, need to be explored further.

Regardless of the physical mechanism, the relative independence of position and quenching greatly simplifies the design of probes for the 5′ nuclease PCR assay. There are three main factors that determine the performance of a double-labeled fluorescent probe in the 5′ nuclease PCR assay. The first factor is the degree of quenching observed in the intact probe. This is characterized by the value of RQ^-, which is the ratio of reporter to quencher fluorescent emissions for a no-template control PCR. Influences on the value of RQ^- include the particular reporter and quencher

TABLE 3 Comparison of Fluorescence Emissions of Single-stranded and Double-stranded Fluorogenic Probes

	518 nm		582 nm		RQ	
Probe	ss	ds	ss	ds	ss	ds
A1-7	27.75	68.53	61.08	138.18	0.45	0.50
A1-26	43.31	509.38	53.50	93.86	0.81	5.43
A3-6	16.75	62.88	39.33	165.57	0.43	0.38
A3-24	30.05	578.64	67.72	140.25	0.45	3.21
P2-7	35.02	70.13	54.63	121.09	0.64	0.58
P2-27	39.89	320.47	65.10	61.13	0.61	5.25
P5-10	27.34	144.85	61.95	165.54	0.44	0.87
P5-28	35.65	462.29	72.39	104.61	0.46	4.43

(ss) Single-stranded. The fluorescence emissions at 518 or 582 nm for solutions containing a final concentration of 50 nM indicated probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 10 mM $MgCl_2$. (ds) Double-stranded. The solutions contained, in addition, 100 nM A1C for probes A1-7 and A1-26, 100 nM A3C for probes A3-6 and A3-24, 100 nM P2C for probes P2-7 and P2-27, or 100 nM P5C for probes P5-10 and P5-28. Before the addition of $MgCl_2$, 120 µl of each sample was heated at 95°C for 5 min. Following the addition of 80 µl of 25 mM $MgCl_2$, each sample was allowed to cool to room temperature and the fluorescence emissions were measured. Reported values are the average of three determinations.

FIGURE 3 Effect of Mg^{2+} concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of $MgCl_2$. The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs. $MgCl_2$ concentration (mM Mg). The key (*upper right*) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe T_m, presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.[1]

The rise in RQ^- values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3′ end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3′ end (A1-26). This is understandable, as the conformation of the 3′ end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3′ end is freer to adopt conformations close to the 5′ reporter dye than is an internally placed quencher. For the other three sets of probes, the interpretation of RQ^- values is less clear-cut. The A3 probes show the same trend as A1, with the 3′ TAMRA probe having a larger RQ^- than the internal TAMRA probe. For the P2 pair, both probes have about the same RQ^- value. For the P5 probes, the RQ^- for the 3′ probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the RQ^- value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on RQ^-.

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes, ΔRQ is much greater for the 3′ TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3′ end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as ΔRQ does not increase when the quencher is placed closer to the 3′ end. This illustrates the importance of being able to use probes with a quencher on the 3′ end in the 5′ nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3′ to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5′ nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3′ end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base pairing and reduce the T_m of a probe. In fact, a 2°C–3°C reduction in T_m has been observed for two probes with internally attached TAMRAs.[2] This disruptive effect would be minimized by placing the quencher at the 3′ end. Thus, probes with 3′ quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3′ quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.[1] demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the ΔF508 mutant. Their probes had TAMRA attached to the seventh nucleotide from

the 5′ end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3′ TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.[10] describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that requires adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761–3766.
2. Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5′ to 3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276–7280.
3. Lyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260: 778–783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* 2: 55–75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*, pp. 303–339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* 58: 719–726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Interspecies homology of sequences in the introns. *Proc. Natl. Acad. Sci.* 82: 6133–6137.
8. du Breuil, R.M., J.M. Patel, and B.V. Mendelow. 1993. Quantitation of β-actin-specific mRNA transcripts using xeno competitive PCR. *PCR Methods Applic.* 3: 57–59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, R.L. Christensen, and E.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences: Poly-dA and poly-A detection. *Nucleic Acids Res.* 22: 2424–2425.

Received December 20, 1994; accepted in revised form March 6, 1995.

GENOME METHODS

Real Time Quantitative PCR

Christian A. Heid,[1] Junko Stevens,[2] Kenneth J. Livak,[2] and P. Mickey Williams[1,3]

[1]BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080;
[2]Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Fan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genome quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Piatak et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)–PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative that it be used properly for quantitation (Raeymaekers 1995). Many early reports of quantitative PCR and RT–PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Ferre 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT–PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as β-actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)–PCR, has been developed and is used widely for PCR quantitation. QC–PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Piatak et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

[3]Corresponding author.
E-MAIL mickey@gene.com; FAX (415) 225-1411.

6:986–994 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QC–PCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT–PCR analysis: (1) agarose gels, (2) fluorescent labeling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fasco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5′ nuclease assay first described by Holland et al. (1991). The method uses the 5′ nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Livak et al. 1995a,b). One fluorescent dye serves as a reporter [FAM (i.e., 6-carboxyfluorescein)] and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

RESULTS

PCR Product Detection in Real Time

The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid-mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pFBTM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact, fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5′–3′ nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human β-actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed ΔRn (or ΔRQ) using the following equation: $\Delta Rn = (Rn^+) - (Rn^-)$, where Rn^+ = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and Rn^- = emission intensity of re-

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points (ΔRns) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The ΔRn mean value is plotted on the *y*-axis, and time, represented by cycle number, is plotted on the *x*-axis. During the early cycles of the PCR amplification, the ΔRn value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base-line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which

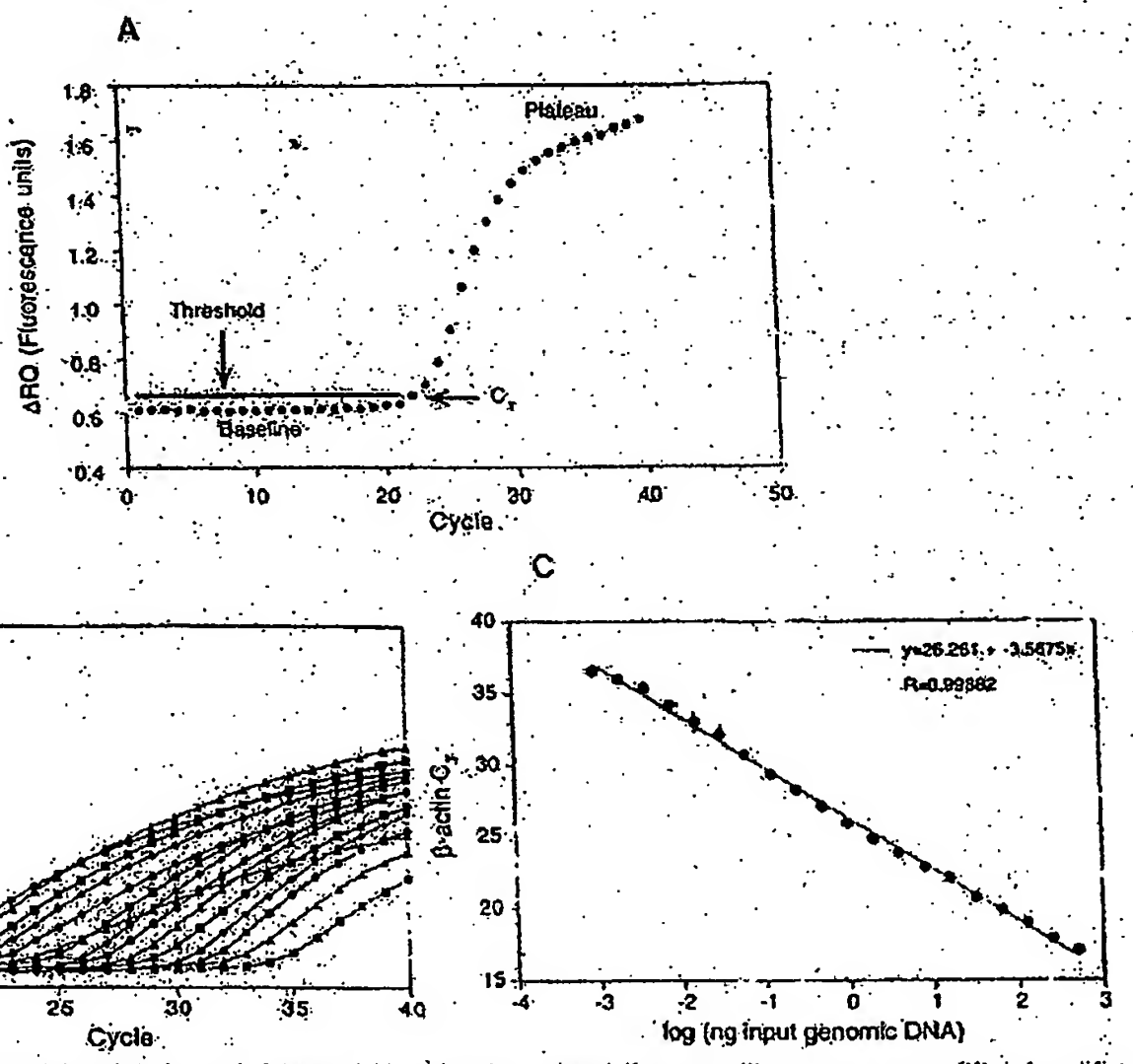


Figure 1 PCR product detection in real time. (*A*) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot. C_T values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (*B*) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with β-actin primers. (*C*) Input DNA concentration of the samples plotted versus C_T. All points represent the mean of triplicate PCR amplifications, and error bars are shown (but not always visible).

the amplification plot crosses the threshold is defined as C_T. C_T is reported as the cycle number at this point. As will be demonstrated, the C_T value is predictive of the quantity of input target.

C_T Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human β-actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the C_T values. Figure 1C represents the C_T values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The C_T values decrease linearly with increasing target quantity. Thus, C_T values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated C_T value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar C_T values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using C_T values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Sequence Detector only spans three logs, resulting in only a 1000-fold dynamic range of input molecules. Thus, C_T values provide accurate measurements over a very large range of relative starting target quantities.

Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing β-actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of C_T values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean C_T values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for β-actin gene quantity. The highest C_T difference between any of the samples was 0.85 and 0.71 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured β-actin C_T value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected C_T value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

Quantitative Analysis of a Plasmid After Transient Transfection

293 cells were transiently transfected with a vec-

Table 1. Reproducibility of Sample Preparation Method

Sample no.	100 ng C_T	mean	standard deviation	CV	25 ng C_T	mean	standard deviation	CV
1	18.24				20.48			
	18.23				20.55			
	18.33	18.27	0.06	0.32	20.5	20.51	0.03	0.17
2	18.33				20.61			
	18.35				20.59			
	18.44	18.37	0.06	0.32	20.41	20.54	0.11	0.54
3	18.3				20.54			
	18.3				20.6			
	18.42	18.34	0.07	0.36	20.49	20.54	0.06	0.28
4	18.15				20.48			
	18.23				20.44			
	18.32	18.23	0.08	0.46	20.38	20.43	0.05	0.26
5	18.4				20.68			
	18.38				20.87			
	18.46	18.42	0.04	0.23	20.63	20.73	0.13	0.61
6	18.54				21.09			
	18.67				21.04			
	19	18.74	0.24	1.26	21.04	21.06	0.03	0.15
7	18.28				20.67			
	18.36				20.73			
	18.52	18.39	0.12	0.66	20.65	20.68	0.04	0.2
8	18.45				20.98			
	18.7				20.84			
	18.73	18.63	0.16	0.83	20.75	20.86	0.12	0.57
9	18.18				20.46			
	18.34				20.54			
	18.36	18.29	0.1	0.55	20.48	20.51	0.07	0.32
10	18.42				20.79			
	18.57				20.78			
	18.66	18.55	0.12	0.65	20.62	20.73	0.1	0.46
Mean	(1–10)	18.42	0.17	0.90		20.66	0.19	0.94

tor containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 μg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β-Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β-actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β-actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β-actin C_T values of the triplicates were plotted (error bars represent one standard deviation). The highest difference between any two sample means was 0.95 C_T. Ten nanograms of total DNA of each sample were also examined for β-actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β-actin C_T value difference was 1.0. As Figure 3 shows, the rate of β-actin C_T change between the 100- and 10-ng samples was similar (slope values range between −3.56 and −3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished

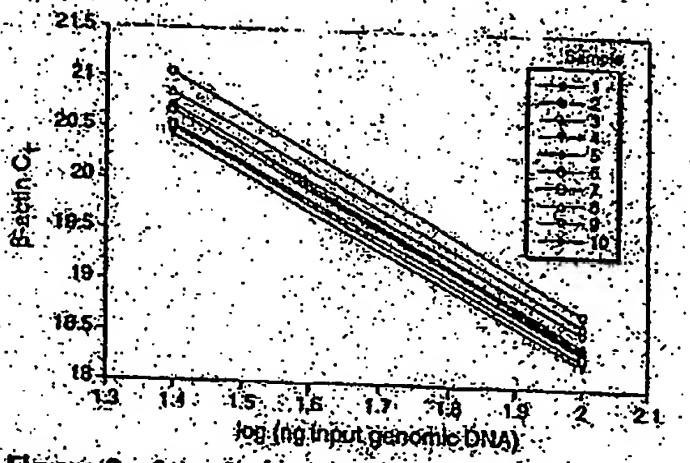


Figure 2. Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs. C_T. In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean β-actin C_T value obtained for each 100-ng sample on a β-actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample, *a*, was obtained by extrapolation to the *x*-axis.

Figure 4A shows the measured (i.e., non-normalized) quantities of factor VIII plasmid DNA (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate

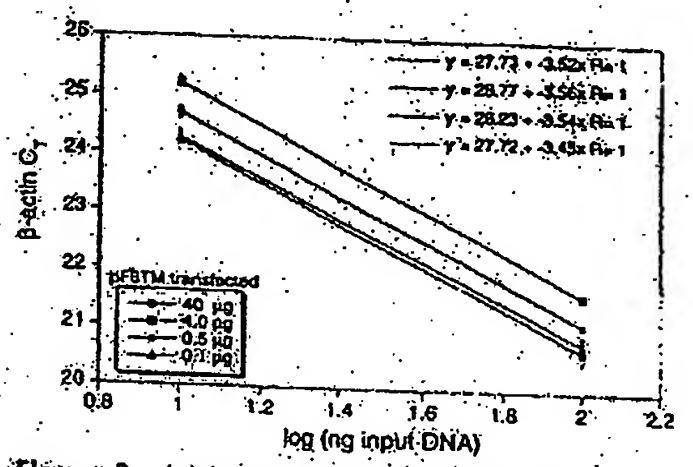


Figure 3. Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1 µg of pF8TM) were analyzed for the β-actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the β-actin C_T values are plotted versus the total input DNA concentration.

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean C_T values increase) with decreasing amounts of plasmid transfected. The mean C_T values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM, *b*, found in each of the four transfections was determined by extrapolation to the *x*-axis of the standard curve in Figure 4B. These uncorrected values, *b*, for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100\ \text{ng}}{a} = \text{actual pF8TM copies per 100 ng of genomic DNA}$$

where *a* = actual genomic DNA in a sample and *b* = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40 µg of plasmid, was 35 pg per 100 ng genomic DNA. This results in ~520 plasmid copies per cell.

DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT-PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e., β-actin) or a "housekeeping" gene for RT-PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

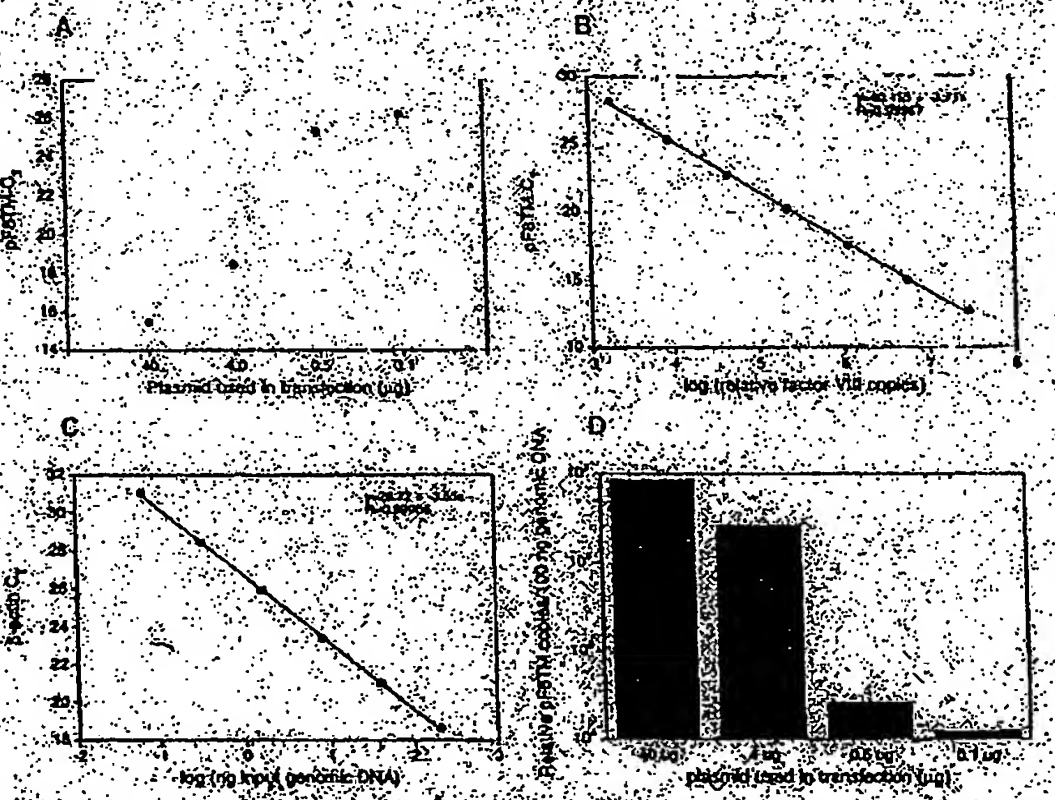


Figure 4 Quantitative analysis of pF8TM in transfected cells. (*A*) Amount of plasmid DNA used for the transfection plotted against the mean C_T value determined for pF8TM remaining 24 hr after transfection. (*B*,*C*) Standard curves of pF8TM and β-actin, respectively. pF8TM DNA (*B*) and genomic DNA (*C*) were diluted serially 1:5 before amplification with the appropriate primers. The β-actin standard curve was used to normalize the results of *A* to 100 ng of genomic DNA. (*D*) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e., β-actin) for quantitative PCR or housekeeping genes for quantitative RT–PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multigene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values, C_T, correlate linearly with relative DNA copy numbers. Real-time quantitative RT–PCR methodology (Gibson et al., this issue) has also been developed. Finally, real-time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications (quantitative gene expression (RT–PCR), gene copy assays (Her2, HIV, etc.), genotyping (knockout mouse analysis), and immuno-PCR).

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes—greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

METHODS

Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol-B from Tel-Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8c2SD (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT–PCR [GeneAmp EZ rTth RNA PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)] using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with *Bam*HI and *Hin*dIII restriction site sequences at the 5′ end) and cloned into pGEM-3Z (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5′-CCCGTGCCAAGAGTGACGTGTC-3′ and F8rev 5′-AAACCTCAGCCTGGATGGTAGG-3′. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 5′ untranslated region of the parent pCIS2.8c2SD plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human β-actin gene was amplified with the primers β-actin forward primer 5′-TCACCCACACTGTGCCCATCTACGA-3′ and β-actin reverse primer 5′-CAGCGGAACCGCTCATTGCCAATGG-3′. The reaction produced a 295-bp PCR product.

Amplification reactions (50 µl) contained a DNA sample, 10× PCR Buffer II (5 µl), 200 µM dATP, dCTP, dGTP, and 400 µM dUTP, 4 mM $MgCl_2$, 1.25 Units AmpliTaq DNA polymerase, 0.5 unit AmpErase uracil *N*-glycosylase (UNG), 50 pmole of each factor VIII primer, and 15 pmole of each β-actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5′-(FAM)AGCTCTCCACCTGCTTCTTCTGTGCCTT(TAMRA)p-3′ and β-actin probe 5′-(FAM)ATGCCCX(TAMRA)CCCCCATGCCATCp-3′ where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801-0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector instrument manual. Briefly, probe T_m should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with 40 cycles of 95°C for 0.5 min and 60°C for 2 min. All reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

REAL TIME QUANTITATIVE PCR

Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM-5 F12 without GHT, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin–streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1 µg were added to 1.5 ml of a solution containing 0.125 M $CaCl_2$ and 1× HEPES. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO_2 for 24 hr, washed with PBS, and resuspended in PBS. The resuspended cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200 µl of 20 mM Tris-HCl at pH 8.0.

ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Bässler, H.A., S.J. Flood, K.J. Livak, J. Marmaro, R. Knorr, and C.A. Batt. 1995. Use of a fluorogenic probe in a PCR-based assay for the detection of Listeria monocytogenes. *App. Environ. Microbiol.* 61: 3724–3728.

Becker-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Meth. Mol. Cell. Biol.* 2: 189–201.

Clementi, M., S. Menzo, P. Bagnarelli, A. Manzin, A. Valenza, and P.E. Varaldo. 1993. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 191–196.

Connor, R.I., H. Mohri, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772–1777.

Eaton, D.L., W.I. Wood, D. Eaton, P.E. Hass, P. Hollingshead, K. Wion, J. Mather, P.M. Lawn, G.A.

Vehar, and C. Gorman. 1986. Construction and characterization of an active factor VIII variant lacking the central one-third of the molecule. *Biochemistry* 25: 8343–8347.

Fasco, M.J., C.P. Treanor, S. Spivack, H.L. Figge, and L.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140–147.

Ferre, F. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Applic.* 2: 1–9.

Furtado, M.R., L.A. Kingsley, and S.M. Wolinsky. 1995. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 69: 2092–2100.

Gibson, U.E.M., C.A. Held, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).

Gorman, C.M., D.R. Gies, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3–10.

Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* 10: 413–417.

Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5′—3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276–7280.

Huang, S.K., H.Q. Xiao, T.J. Kleine, G. Paciotti, D.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688–2694.

Huang, S.K., M. Yi, E. Palmer, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen, Amb a 5. *J. Immun.* 154: 6157–6162.

Kellogg, D.E., J.J. Sninsky, and S. Kowk. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202–208.

Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761–3766.

Livak, K.J., S.J. Flood, J. Marmaro, W. Giusti, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic.* 4: 357–362.

Livak, K.J., J. Marmaro, and J.A. Todd. 1995b. Towards fully automated genome-wide polymorphism screening [Letter]. *Nature Genet.* 9: 341–342.

Mulder, J., N. McKinney, C. Christopherson, J. Sninsky, L. Greenfield, and S. Kwok. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292–300.

Pang, S., Y. Koyanagi, S. Miles, C. Wiley, H.V. Vinters, and I.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85–89.

Piatak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1993a. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *BioTechniques* 14: 70–81.

Piatak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hahn, G.M. Shaw, and J.D. Lifson. 1993b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR [see Comments]. *Science* 259: 1749–1754.

Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495–503.

Raeymaekers, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 91–94.

Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750–768.

Slamon, D.J., G.M. Clark, S.G. Wong, W.J. Levin, A. Ullrich, and W.L. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177–182.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503–517.

Tan, X., X. Sun, C.F. Gonzalez, and W. Hsueh. 1994. PAF and TNF increase the precursor of NF-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1215: 157–162.

Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201–5205.

Williams, S., C. Schwer, A. Krishnarao, C. Held, B. Karger, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (in press).

Received June 3, 1996; accepted in revised form July 29, 1996.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14717–14722, December 1998
Cell Biology, Medical Sciences

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE‖, ROBERT L. COHEN*, MONA F. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,‡‡, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, ‡Molecular Biology, ‖Scientific Computing, and ¶Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT **Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (*i*) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracyline repressible promoter, and (*ii*) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.**

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3β (GSK-3β) resulting in an increase in β-catenin levels. Stabilized β-catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β-catenin levels (9). APC is phosphorylated by GSK-3β, binds to β-catenin, and facilitates its degradation. Mutations in either APC or β-catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)-β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies c-*myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

cDNA was synthesized from 2 μg of poly(A)$^+$ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μg of poly(A)$^+$ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λgt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λgt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μM of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

***In Situ* Hybridization.** ^{33}P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNAzol.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISP*s and c-*myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{(\Delta ct)}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The ∂-method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin–Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription–PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1 *A* and *B*). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β-catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription–PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISP*s may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of ≈40,000 (M_r 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2*A*).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of ≈27,000 (M_r 27 K) (Fig. 2*B*). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

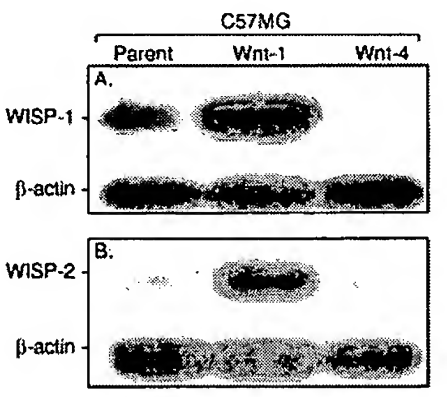


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (*A*) and *WISP-2* (*B*) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)$^+$ RNA (2 μg) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3′ untranslated region. Blots were rehybridized with human β-actin probe.

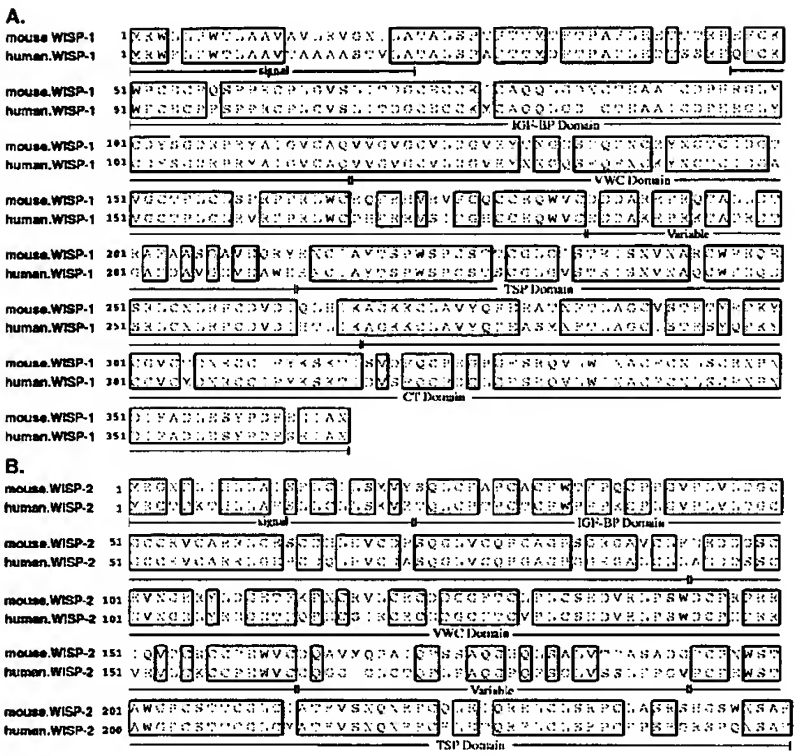


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (*A*) and mouse and human *WISP-2* (*B*). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the WISP-1 protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called WISP-3. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. WISP-3 has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human WISP proteins shows that WISP-1 and WISP-3 are the most similar (42% identity), whereas WISP-2 has 37% identity with WISP-1 and 32% identity with WISP-3 (Fig. 3*A*).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF-β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3*B*) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGCCXXC) conserved in most insulin-like growth factor (IGF)-

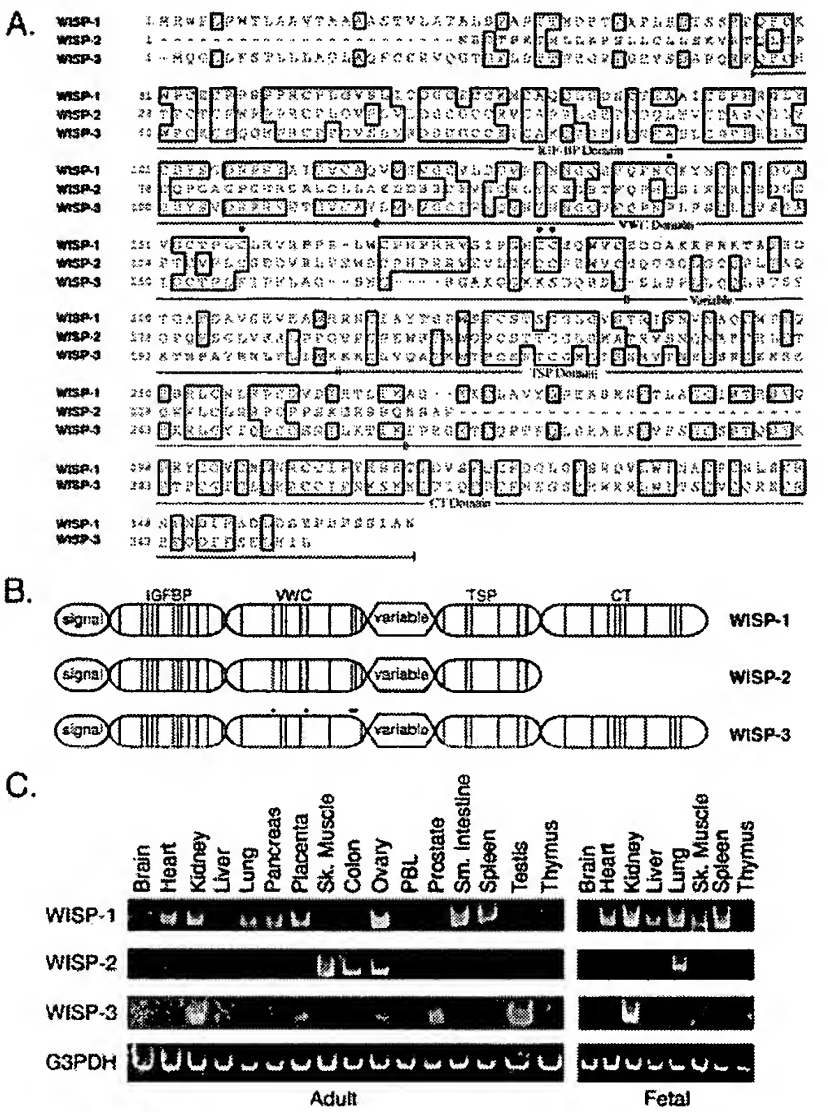


FIG. 3. (*A*) Encoded amino acid sequence alignment of human WISPs. The cysteine residues of WISP-1 and WISP-2 that are not present in WISP-3 are indicated with a dot. (*B*) Schematic representation of the WISP proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in WISP-3 are indicated with a dot. (*C*) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in WISP-2 and WISP-3, whereas WISP-1 has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of WISP-3 differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3 *A* and *B*). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCSxxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in WISP-2 (Fig. 3 *A* and *B*). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that WISPs are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISP*s was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3*C*). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

***In Situ* Localization of *WISP-1* and *WISP-2*.** Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 *A–D*). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 *E–H*). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

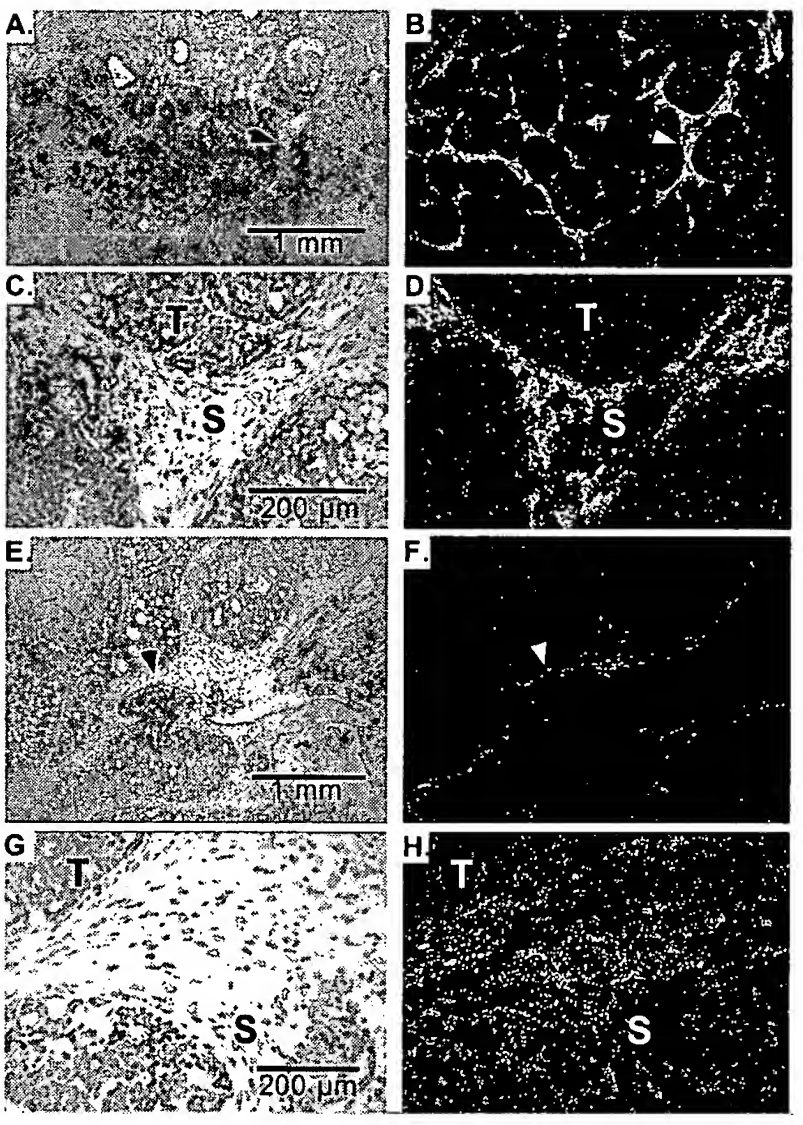


FIG. 4. (*A*, *C*, *E*, and *G*) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in *B* and *D*. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (*A* and *B*), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (*C* and *D*), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in *E–H*. At low power (*E* and *F*), expression of WISP-2 is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (*G* and *H*).

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *nov*H family member (27) and roughly 4 Mbs distal to c-*myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, c-*myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as c-*myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the c-*myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 *A* and *B*). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for c-*myc*, indicating that the c-*myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

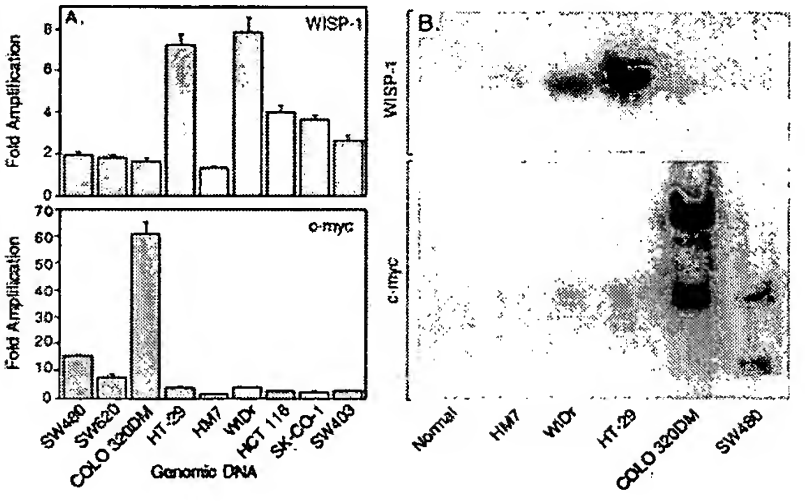


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (*A*) Amplification in cell line DNA was determined by quantitative PCR. (*B*) Southern blots containing genomic DNA (10 μg) digested with *Eco*RI (*WISP-1*) or *Xba*I (c-*myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human c-*myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

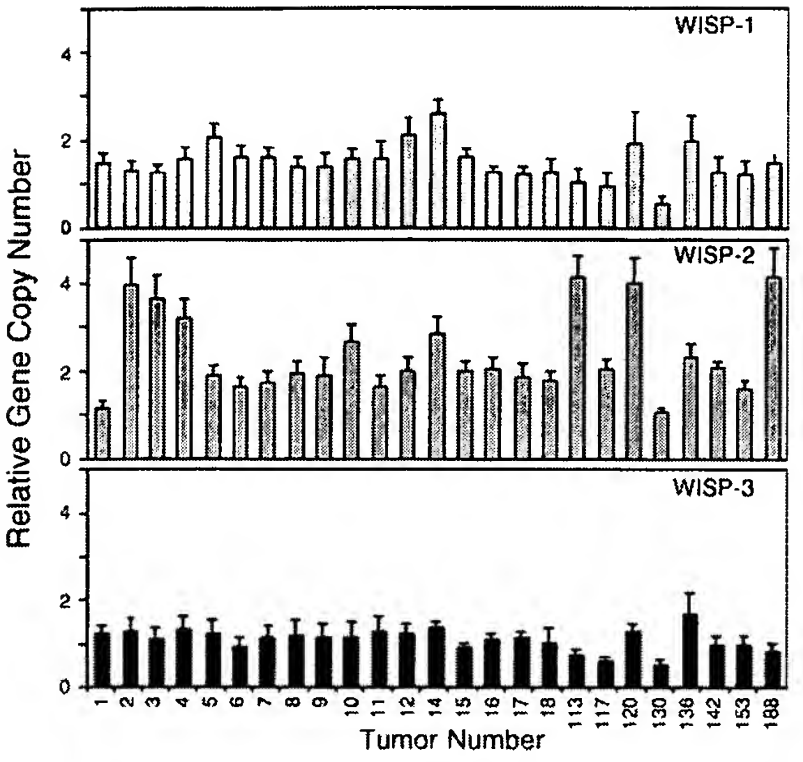


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means ± SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

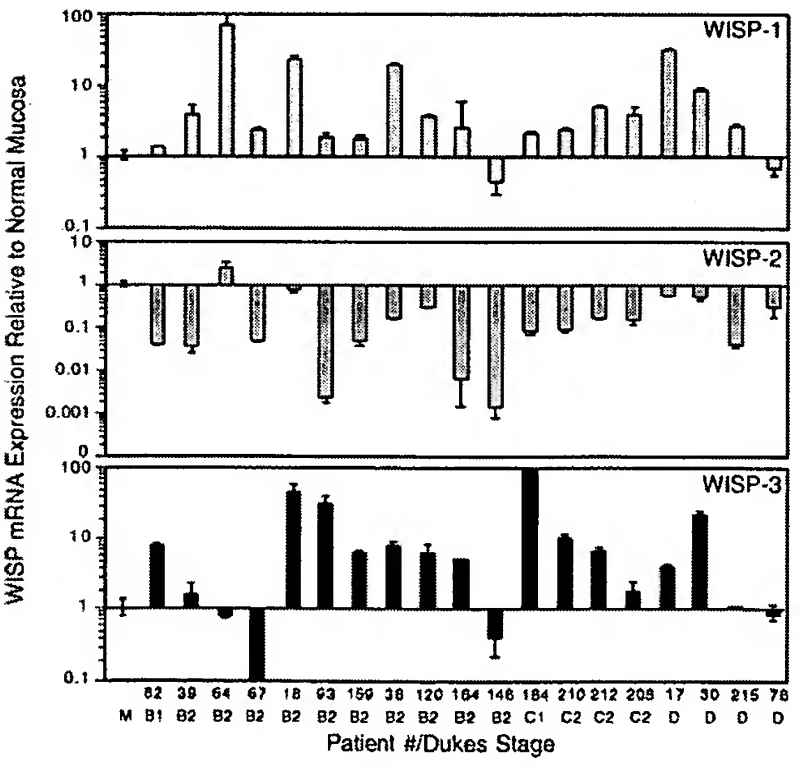


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means ± SEM from one experiment done in triplicate. The experiment was repeated at least twice.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracyline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISP*s in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISP*s are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β-catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β-catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISP*s.

The *WISP*s define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of WISP-2 is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, WISP-1, and WISP-3. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF-β, platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that WISP-1 and WISP-3 may exist as dimers, whereas WISP-2 exists as a monomer. If the CT domain is also important for receptor binding, WISP-2 may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF-β1, which is the stimulus for stromal proliferation (34). TGF-β1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and WISPs in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β-catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β-catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISP*s. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISP*s as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISP*s in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

1. Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* **11,** 3286–3305.
2. Dale, T. C. (1998) *Biochem. J.* **329,** 209–223.
3. Nusse, R. & Varmus, H. E. (1982) *Cell* **31,** 99–109.
4. van Ooyen, A. & Nusse, R. (1984) *Cell* **39,** 233–240.
5. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* **55,** 619–625.
6. Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell. Biol.* **10,** 182–187.
7. Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* **86,** 391–399.
8. Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell. Biol.* **18,** 1248–1256.
9. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3046–3050.
10. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* **281,** 1509–1512.
11. Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6025–6030.
12. Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* **46,** 1001–1009.
13. Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell. Biol.* **14,** 6278–6286.
14. Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* **8,** 1349–1358.
15. Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* **187,** 289–296.
16. Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell. Biol.* **18,** 6131–6141.
17. Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* **8,** 171–179.
18. Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell. Biol.* **16,** 1326–1334.
19. Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6355–6360.
20. Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* **9,** 2729–2732.
21. Bork, P. (1993) *FEBS Lett.* **327,** 125–130.
22. Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12981–12986.
23. Joliot, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell. Biol.* **12,** 10–21.
24. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* **264,** 19514–19527.
25. Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* **265,** 2852–2855.
26. Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell. Biol.* **113,** 195–205.
27. Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* **7,** 2529–2534.
28. Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell. Genet.* **57,** 109–111.
29. Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* **1,** 88–94.
30. Garte, S. J. (1993) *Crit. Rev. Oncog.* **4,** 435–449.
31. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* **276,** 1268–1272.
32. Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 269–291.
33. Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* **273,** 3090–3096.
34. Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell. Biol.* **29,** 153–161.
35. Wernert, N. (1997) *Virchows Arch.* **430,** 433–443.
36. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* **54,** 4257–4260.
37. Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* **6,** 187–194.
38. Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., *et al.* (1998) *EMBO J.* **17,** 3052–3065.
39. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* **275,** 1787–1790.
40. Lu, L. H. & Gillett, N. (1994) *Cell Vision* **1,** 169–176.

methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described[20,21]. Briefly, after antigen pulsing (30 μg ml^{-1} TTCF) with tetrapeptides (1–2 mg ml^{-1}), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with 1 μCi of ^{3}H-thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating 200 μg TTCF with 0.25 μg pig kidney legumain in 500 μl 50 mM citrate buffer, pH 5.5, for 1 h at 37 °C.

Glycopeptide digestions. The peptides HIDNEEDI, HIDN(*N*-glucosamine)EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQHLFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQHLFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography[11]. Glycopeptides corresponding to residues 622–642 and 421–452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at 30 °C with 5–50 mU ml^{-1} pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of 10 mg ml^{-1} α-cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 5 November 1998.

1. Chen, J. M. *et al.* Cloning, isolation, and characterisation of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090–8098 (1997).
2. Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 305, 208–213 (1993).
3. Dalton, J. P., Hola Jamriska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575–580 (1995).
4. Bennett, K. *et al.* Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1519–1524 (1992).
5. Riese, R. J. *et al.* Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357–366 (1996).
6. Rodriguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894–2898 (1992).
7. Hewitt, E. W. *et al.* Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693–4699 (1997).
8. Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821–850 (1997).
9. Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93–102 (1998).
10. Fiorchi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377–382 (1997).
11. Lu, J. & van Halbeek, H. Complete ^{1}H and ^{13}C resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1–21 (1996).
12. Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50–54 (1996).
13. Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295–298 (1997).
14. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705–711 (1998).
15. Botarelli, P. *et al.* N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128–3132 (1991).
16. Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101–4109 (1990).
17. Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535–559 (1981).
18. Makoff, A. J., Ballantine, S. P., Smellwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043–1046 (1989).
19. Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
20. Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537–539 (1985).
21. Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543–553 (1997).

Acknowledgements. We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte-preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti*†, Scot A. Marsters*†, David A. Lawrence*†, Margaret Roy*, Frank C. Kischkel*, Patrick Dowd*, Arthur Huang*, Christopher J. Donahue*, Steven W. Sherwood*, Daryl T. Baldwin*, Paul J. Godowski*, William I. Wood*, Austin L. Gurney*, Kenneth J. Hillan*, Robert L. Cohen*, Audrey D. Goddard*, David Botstein‡ & Avi Ashkenazi*

**Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA*
‡Department of Genetics, Stanford University, Stanford, California 94305, USA
† These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells[1]. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (DcR3), that binds to FasL and inhibits FasL-induced apoptosis. The DcR3 gene was amplified in about half of 35 primary lung and colon tumours studied, and DcR3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily[2]. Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (DcR3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino-terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)[3], DcR3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-asscociated, molecule. We expressed a recombinant, histidine-tagged form of DcR3 in mammalian cells; DcR3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). DcR3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of DcR3 and OPG are conserved; however, the carboxy-terminal portion of DcR3 is 101 residues shorter.

We analysed expression of DcR3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high DcR3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of DcR3, we generated a recombinant, Fc-tagged DcR3 protein. We tested binding of DcR3–Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). DcR3–Fc showed a significant increase in binding to cells transfected with FasL[4] (Fig. 2a), but not to cells transfected with TNF[5], Apo2L/TRAIL[6,7], Apo3L/TWEAK[8,9], or OPGL/TRANCE/

RANKL[10–12] (data not shown). DcR3–Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3–Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3–Fc and Fas–Fc bound to soluble FasL with a comparable affinity ($K_d = 0.8 \pm 0.2$ and 1.1 ± 0.1 nM, respectively; Fig. 2e), and that DcR3–Fc could block nearly all of the binding of soluble FasL to Fas–Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3–Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3–Fc and Fas–Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at ~0.1 μg ml^{-1}. Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3–Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process[1]. Consistent with previous results[13], activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas–Fc blocked this effect substantially (Fig. 3c); DcR3–Fc blocked the induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes[1,14–16]. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3–Fc and Fas–Fc each reduced killing of target cells from ~65% to ~30%, with half-maximal inhibition at ~1 μg ml^{-1}; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3–Fc and Fas–Fc concentrations were required to block natural killer cell activity compared with those required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL[17].

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain

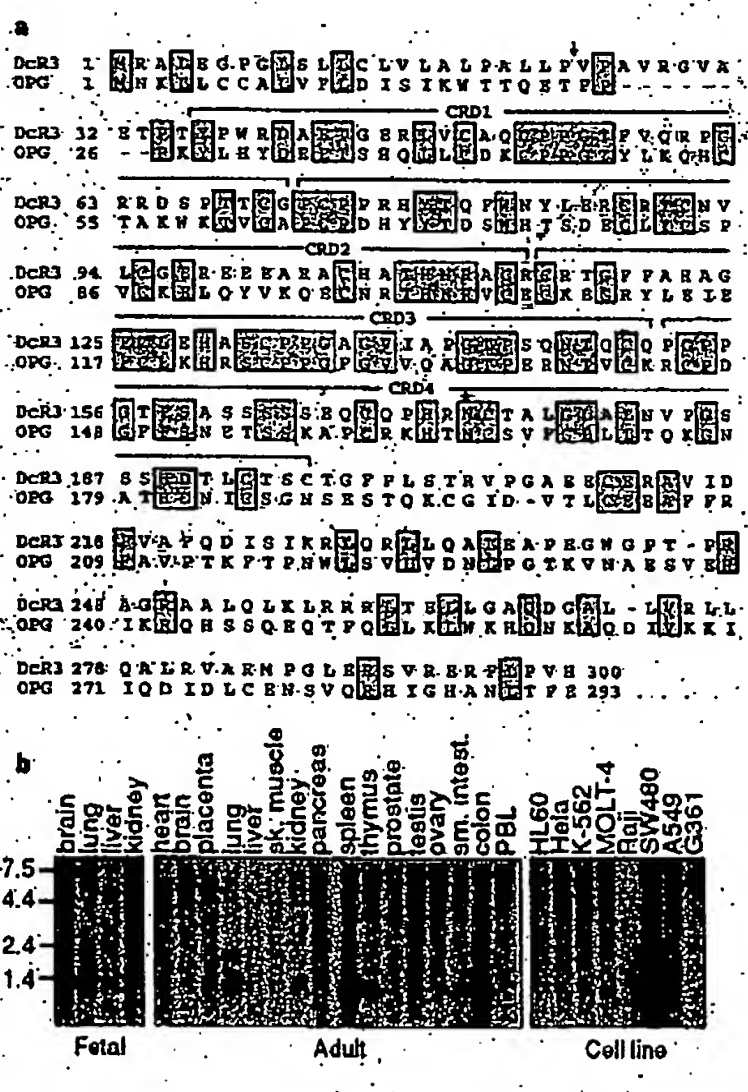


Figure 1 Primary structure and expression of human DcR3. a, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1–4), and the *N*-linked glycosylation site (asterisk) are shown. b, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)$^+$ RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.

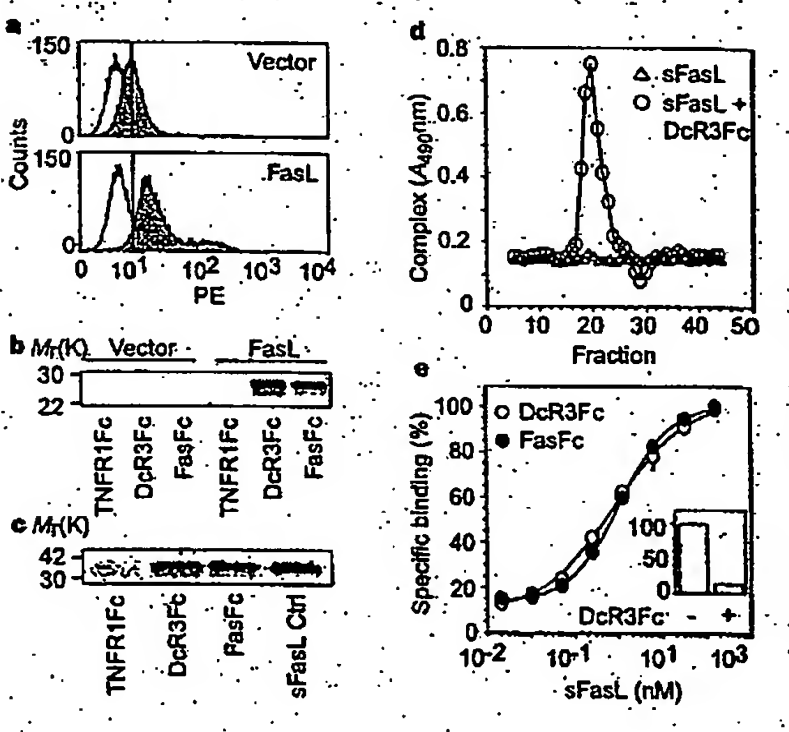


Figure 2 Interaction of DcR3 with FasL. a, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3–Fc (solid line, shaded area), TNFR1–Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ($P < 0.001$) between the binding of DcR3–Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. b, 293 cells were transfected as in a and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. c, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1–Fc, DcR3–Fc or Fas–Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. d, Flag-tagged sFasL was incubated with DcR3–Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3–Fc and sFasL–Flag. e, Equilibrium binding of DcR3–Fc or Fas–Fc to sFasL–Flag. Inset, competition of DcR3–Fc with Fas–Fc for binding to sFasL–Flag.

reaction (PCR)[18] in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this, we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epicentre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG[2,19].

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas[20]. A second mechanism involves proteolytic shedding of FasL from the cell surface[17]. DcR3 competes with Fas for

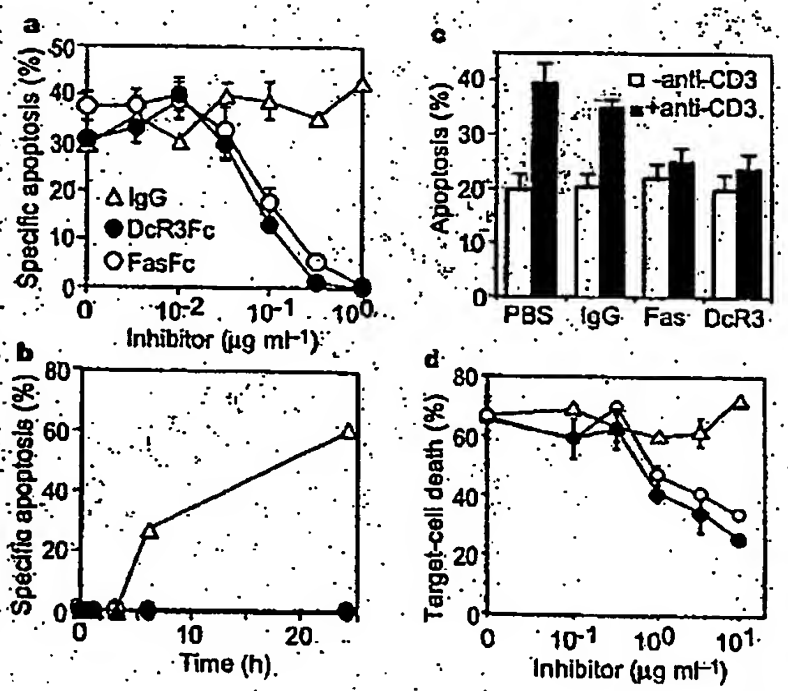


Figure 3 Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml^{-1}) oligomerized with anti-Flag antibody (0.1 μg ml^{-1}) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL–Flag plus anti-Flag antibody as in a, in presence of 1 μg ml^{-1} DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml^{-1}). After 16 h, apoptosis of CD4$^+$ cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with ^{51}Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of ^{51}Cr (mean ± s.d. for two donors, each in triplicate).

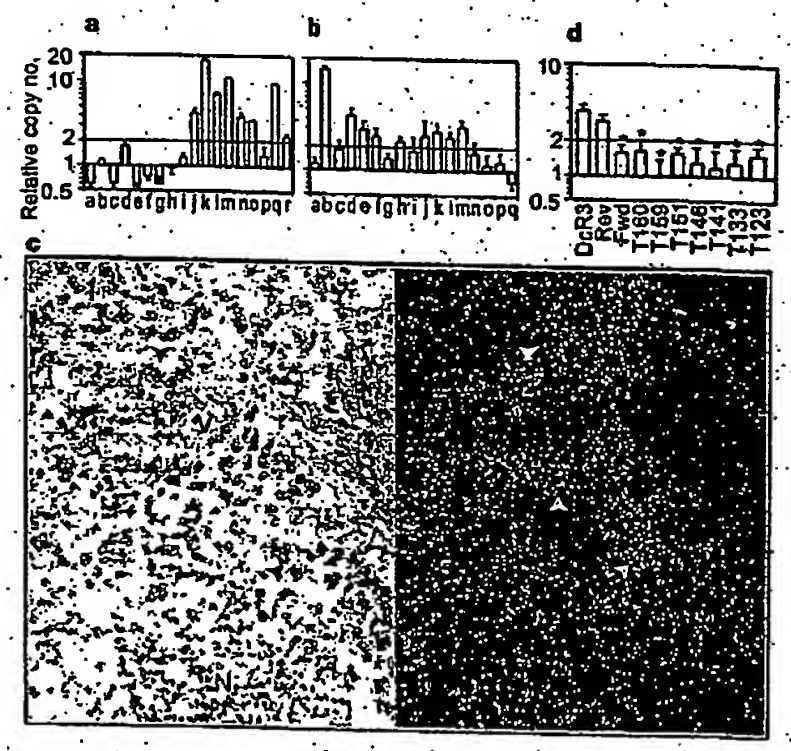


Figure 4 Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (l), and one bronchial adenocarcinoma (i). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (**b**). Data are from two experiments done in duplicate. Asterisk indicates $P < 0.01$ for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described[21]. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L[22]. Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG[3], which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L[19]. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response[2]. Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

Methods

Isolation of DcR3 cDNA. Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in Lifeseq™ (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

Fc-fusion proteins (immunoadhesins). The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described[23].

Fluorescence-activated cell sorting (FACS) analysis. We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL[4] (2 μg), together with pRK5 encoding CrmA (2 μg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3–Fc or TNFR1–Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov–Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3–Fc, as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

Immunoprecipitation. Human 293 cells were transfected as above, and metabolically labelled with [^{35}S]cysteine and [^{35}S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 μM), the medium was immunoprecipitated with DcR3–Fc, Fas–Fc or TNFR1–Fc (5 μg), followed by protein A–Sepharose (Repligen). The precipitates were resolved by SDS–PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 μg) (Alexis) was incubated with each Fc-fusion protein (1 μg), precipitated with protein A–Sepharose, resolved by SDS–PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

Analysis of complex formation. Flag-tagged soluble FasL (25 μg) was incubated with buffer or with DcR3–Fc (40 μg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3–Fc–FasL complex in each fraction was analysed by placing 100 μl aliquots into microtitre wells precoated with anti-human IgG (Boehringer) to capture DcR3–Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin–horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3–Fc homodimers to two soluble FasL homotrimers.

Equilibrium binding analysis. Microtitre wells were coated with anti-human IgG, blocked with 2% BSA in PBS. DcR3–Fc or Fas–Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas–Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3–Fc.

T-cell AICD. CD3$^+$ lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotech), stimulated with phytohaemagglutinin (PHA; 2 μg ml^{-1}) for 24 h, and cultured in the presence of interleukin-2 (100 U ml^{-1}) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4$^+$ cells[24].

Natural killer cell activity. Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotech), and incubated for 16 h with ^{51}Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3–Fc, Fas–Fc or human IgG1. Target-cell death was determined by release of ^{51}Cr in effector–target co-cultures relative to release of ^{51}Cr by detergent lysis of equal numbers of Jurkat cells.

Gene-amplification analysis. Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR[18] using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5′-CTTCTTCGCGCACGCTG-3′ and 5′-ATCACGCCGGCACCAG-3′ and the fluorogenic probe sequence was 5′-(FAM-ACACGATGCGTGCTCCAAGCAG AAp-(TAMARA), where FAM is 5′-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula $2^{(\Delta CT)}$, where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

1. Nagata, S. Apoptosis by death factor. *Cell* 88, 355–365 (1997).
2. Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959–962 (1994).
3. Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309–319 (1997).
4. Suda, T., Takahashi, T., Goltstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169–1178 (1993).
5. Pennica, D. *et al.* Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724–729 (1984).
6. Pitti, R. M. *et al.* Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687–12690 (1996).
7. Wiley, S. R. *et al.* Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673–682 (1995).
8. Marsters, S. A. *et al.* Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525–528 (1998).
9. Chicheportiche, Y. *et al.* TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401–32410 (1997).
10. Wong, B. R. *et al.* TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190–25194 (1997).
11. Anderson, D. M. *et al.* A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175–179 (1997).
12. Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165–176 (1998).
13. Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo-1/(Fas/CD95). *Nature* 373, 438–441 (1995).
14. Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235–1238 (1995).
15. Medvedev, A. E. *et al.* Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394–404 (1997).
16. Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13–18 (1997).
17. Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31–36 (1998).
18. Gelmini, S. *et al.* Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752–758 (1997).
19. Emery, J. G. *et al.* Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363–14367 (1998).
20. Wallach, D. Playing death under control. *Nature* 388, 123–125 (1997).
21. Colotta, F. *et al.* Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* 261, 472–475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* 281, 1305–1308 (1998).
23. Ashkenazi, A. & Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* 9, 195–200 (1997).
24. Marsters, S. *et al.* Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

Crystal structure of the ATP-binding subunit of an ABC transporter

Li-Wei Hung*, Iris Xiaoyan Wang†, Kishiko Nikaido†, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim*‡

**E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA*

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes[1]. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters[2]. Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains[1]. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*[1,3–5] is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, $HisQMP_2$, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins[5], is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM[6]. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis[5], the requirement for both subunits to be present for activity[5], and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer[3]. HisP has been purified and characterized in an active soluble form[3] which can be reconstituted into a fully active membrane-bound complex[5].

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded β-sheet (β3 and β8–β12) spans both arms of the L, with a domain of α- plus β-type structure (β1, β2, β4–β7, α1 and α2) on one side (within arm I) and a domain of mostly α-helices (α3–α9) on the

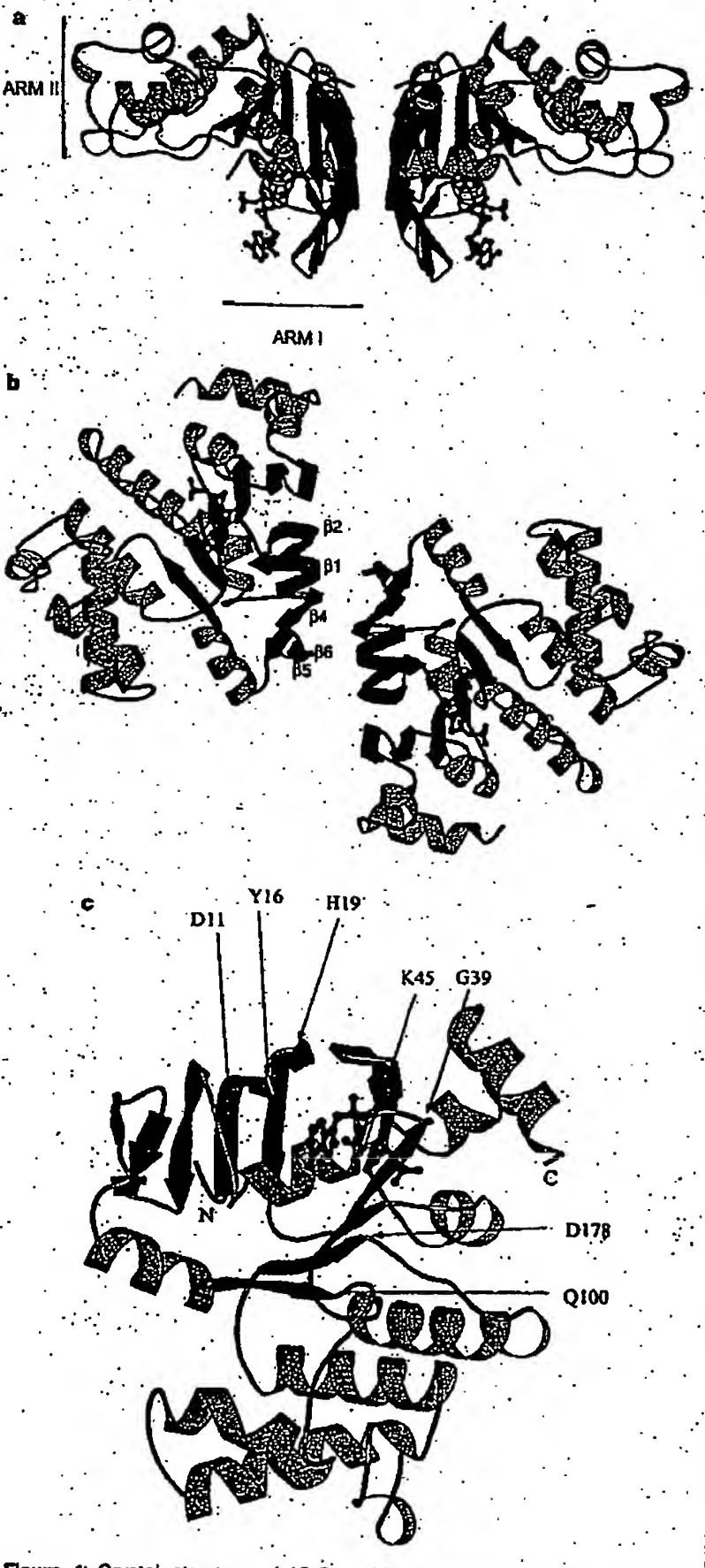


Figure 1 Crystal structure of HisP. a, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane. α-Helices are shown in orange and β-sheets in green. b, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in a. The β-strands at the dimer interface are labelled. c, View of one monomer from the bottom of arm I, as shown in a, towards arm II, showing the ATP-binding pocket. a–c, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in c. These figures were prepared with MOLSCRIPT[29]. N, amino terminus; C, C terminus.

Int. J. Cancer: 78, 661–666 (1998)

Publication of the International Union Against Cancer
Publication de l'Union Internationale Contre le Cancer

NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE[1,2], Martine OLIVI[1], Marie-Hélène CHAMPÈME[2], Dominique VIDAUD[1], Rosette LIDEREAU[2] and Michel VIDAUD[1]*

[1]*Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France.*
[2]*Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France*

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erb*B2) in breast tumors. Extra copies of *myc*, *ccnd1* and *erb*B2 were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNA; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.
© *1998 Wiley-Liss, Inc.*

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmins) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erb*B2 (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erb*B2 proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuuring *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy surrounding the link suggested by Slamon *et al.* (1987) between *erb*B2 amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5′ nuclease assay first described by Holland *et al.* (1991). The latter uses the 5′ nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5′ end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3′ end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris; 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-4407-1754.
E-mail: mvidaud@teaser.fr

Received 2 May 1998; Revised 30 June 1998.

cycle, the fluorescent hybridization probe is hydrolyzed by the 5′-3′ nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (*i.e.*, ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the Rn (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): *(i)* the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridation of both the primers and the probe is necessary to generate a signal; *(ii)* the C_t (threshold cycle) value used for quantification is measured when PCR amplification is still in the log-phase of PCR product accumulation. This is the main reason why C_t is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; *(iii)* use of C_t values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; *(iv)* The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; *(v)* the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; *(vi)* the assay is rapid, as results are available just one minute after thermal cycling is complete; *(vii)* the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erb*B2), as well as 2 genes (*alb* and *app*), located in a chromosome region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

MATERIAL AND METHODS

Tumor and blood samples.

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguenin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

Real-time PCR

Theoretical basis.. Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter C_t (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring C_t and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (*i.e.*, lack of extensive degradation) are both difficult to assess. We therefore also quantified a control gene *(alb)* mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast-tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (}\textit{app, myc, ccnd1, erb}\text{B2)}}{\text{copy number of reference gene (}\textit{alb}\text{)}}$$

Primers, probes, reference human genomic DNA and PCR consumables. Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAgency (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

Standard-curve construction. The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatak *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophorezed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/μl. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from 10^{-7} (10^5 copies of each gene) to 10^{-10} (10^2 copies). This series of diluted PCR products was aliquoted and stored at −80°C until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

PCR amplification. Amplification mixes (50 μl) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes), 10× TaqMan buffer (5 μl), 200 μM dATP, dCTP, dGTP, and 400 μM dUTP, 5 mM $MgCl_2$, 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at 50°C and 10 min at 95°C. Thermal cycling consisted of 40 cycles at 95°C for 15 s and 65°C for 1 min. Each assay included: a standard curve (from 10^5 to 10^2 copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

Equipment for real-time detection. The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates C_t and determines the starting copy number in the samples.

Determination of gene amplification. Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

RESULTS

To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erb*B2 proto-oncogenes, and the β-amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/μl. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as 10^2 copies or as many as 10^5 copies.

*Copy-number ratio of the 2 reference genes (*app *and* alb*)*

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA

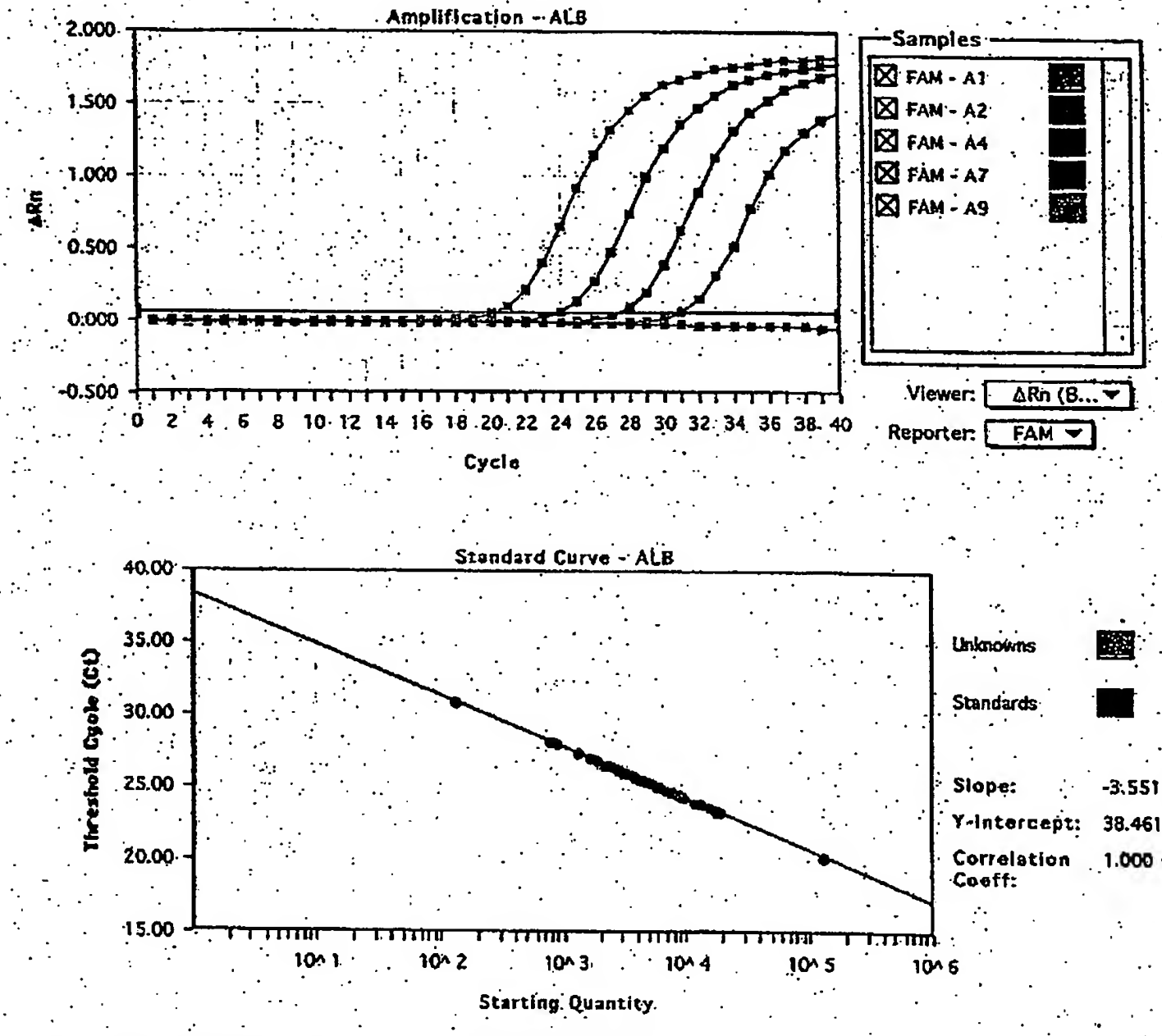


FIGURE 1 – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from 10^5 (A9), 10^4 (A7), 10^3 (A4) to 10^2 (A2) and a no-template control (A1). Cycle number is plotted *vs.* change in normalized reporter signal (ΔRn). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal (Rn). ΔRn represents the normalized reporter signal (Rn) minus the baseline signal established in the first 15 PCR cycles. ΔRn increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau. C_t (threshold cycle) represents the fractional cycle number at which a significant increase in Rn above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number *vs.* C_t (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.

samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean 1.02 ± 0.21), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6, mean 1.06 ± 0.25), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

myc, ccnd1 and erbB2 gene dose in normal leukocyte DNA

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean 0.84 ± 0.22) for *myc*, 0.7 to 1.6 (mean 1.06 ± 0.23) for *ccnd1* and 0.6 to 1.3 (mean 0.91 ± 0.19) for *erb*B2. Since N values for *myc*, *ccnd1* and *erb*B2 in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

myc, ccnd1 and erbB2 gene dose in breast-tumor DNA

myc, *ccnd1* and *erb*B2 gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erb*B2 (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erb*B2, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (TI18). The 3 genes were never found to be co-amplified in the same tumor. *erb*B2 and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erb*B2 in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erb*B2 copy number (N < 0.5), suggesting that they bore deletions of the 17q21 region (the site of *erb*B2). No such decrease in copy number was observed with the other 2 proto-oncogenes.

Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erb*B2 amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers (N ≥ 5). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erb*B2) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

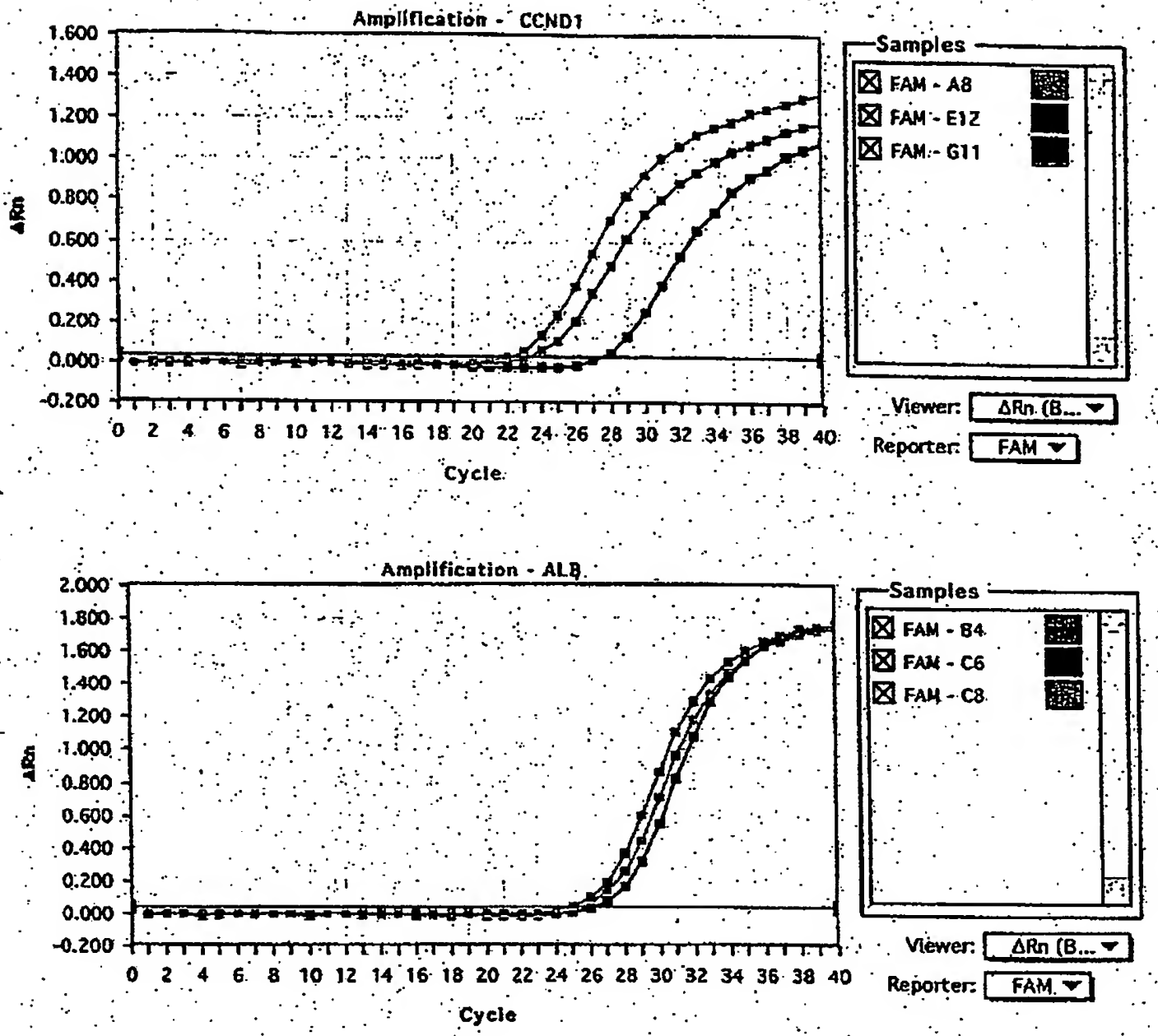
In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on C_t values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI Prism 7700 Sequence Detection System enables C_t to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the C_t value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the C_t ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

The only potential disavantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that is cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (*i*) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (*ii*) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (*iii*) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erb*B2 amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (*iv*) The maxima of *ccnd1* and *erb*B2 over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I – DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erb*B2 GENES IN 108 HUMAN BREAST TUMORS

Gene	Amplification level (N)			
	<0.5	0.5–1.9	2–4.9	≥5
myc	0	97 (89.8%)	11 (10.2%)	0
ccnd1	0	83 (76.9%)	17 (15.7%)	8 (7.4%)
*erb*B2	5 (4.6%)	87 (80.6%)	8 (7.4%)	8 (7.4%)



Tumor	CCND1 C_t	CCND1 Copy number	ALB C_t	ALB Copy number
T118	27.3	4605	26.5	4365
T133	23.2	61659	25.2	10092
T145	22.1	125892	25.6	7762

FIGURE 2 – *ccnd1* and alb gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the C_t of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (*v*) The *erb*B2 copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron *et al.*, 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erb*B2 amplification in a small series of breast tumors (n = 25), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-

TABLE II – EXAMPLES OF *ccnd1* GENE-DOSAGE RESULTS FROM 3 BREAST TUMORS[1]

Tumor	*ccnd1*			*alb*			N*ccnd1/alb*
	Copy number	Mean	SD	Copy number	Mean	SD	
T118	4525			4223			
	4605	4603	77	4365	4325	89	1.06
	4678			4387			
T133	59821			9787			
	61659	61100	1111	10092	10137	375	6.03
	61821			10533			
T145	128563			7321			
	125892	125392	3448	7762	7672	316	16.34
	121722			7933			

[1]For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (N*ccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers (≥5-fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erb*B2) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erb*B2 (but not of the other 2 proto-oncogenes) in several tumors; *erb*B2 is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

REFERENCES

AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEYEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erb*B2 gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, 64, 291–297 (1995).

BERNS, E.M.J.J., KLIJN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTENGEN, H. and FOEKENS, J.A., c-*myc* amplification is a better prognostic factor than HER2/*neu* amplification in primary breast cancer. *Cancer Res.*, 52, 1107–1113 (1992).

BIÈCHE, I. and LIDEREAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, 14, 227–251 (1995).

BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., c-*myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, 51, 687–691 (1992).

CELI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDINER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (gd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, 21, 304–310 (1994).

COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, 57, 4360–4367 (1997).

DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erb*B-2 and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, 40, 271–281 (1996).

GELMINI, S., ORIANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure c-*erB-2* oncogene amplification. *Clin. Chem.*, 43, 752–758 (1997).

GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, 6, 995–1001 (1996).

HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, 6, 986–994 (1996).

HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5′ to 3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, 88, 7276–7280 (1991).

KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, 91, 2156–2160 (1994).

LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, 21, 3761–3766 (1993).

LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, 93, 125–128 (1990).

MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, F., CIRRINCIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., c-*erb*B-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, 330, 1260–1266 (1994).

PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of HER-2/*neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, 13, 63–72 (1996).

PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, 14, 70–80 (1993).

SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILEE, P., CORNELISSE, C.J., VAN DE VIJVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, 52, 5229–5234 (1992).

SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science*, 235, 177–182 (1987).

SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDOVE, J., ULLRICH, A. and PRESS, M.F., Studies of the HER-2/*neu* proto-oncogene in human breast and ovarian cancer. *Science*, 244, 707–712 (1989).

VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying HER-2/*neu* status in breast cancer. *Int. J. Cancer*, 65, 129–133 (1996).

GENES V

Benjamin Lewin

OXFORD UNIVERSITY PRESS
Oxford New York Tokyo
1994

~10×). Also, some tumor lines have amplified *ras* genes. A 20-fold increase in the level of a nontransforming Ras protein is sufficient to allow the transformation of some cells. The effect has not been fully quantitated, but it suggests the general conclusion that oncogenesis depends on overactivity of Ras protein, and is caused either by increasing the amount of protein or (probably more efficiently) by a variety of mutations that increase the activity of the protein.

Transfection by DNA can be used to transform only certain cell types. Although transforming oncogenes have been isolated from both rodent and human cells, most targets for transformation by transfection with oncogenes have been rodent fibroblasts in culture. (In fact, the difference in the source of the oncogene [human] and the recipient cell [rodent] is an important factor in allowing the donor gene to be distinguished unequivocally from recipient DNA.) Limitations of the assay explain why relatively few oncogenes have been detected by transfection. This system has been most effective with *ras* genes, where there is extensive correlation between mutations that activate *c-ras* genes in transfection and the occurrence of tumors.

Insertion, translocation, or amplification may activate proto-oncogenes

A variety of genomic changes can activate proto-oncogenes, sometimes involving a change in the target gene itself, sometimes activating it without changing the protein product. In cases of insertion and translocation, there is evidence that the genomic change is the causative event; in cases of amplification there is a correlation with tumorigenesis, but no direct proof for a causative role.

Many tumor cell lines have visible regions of chromosomal amplification, as shown by homogeneously staining regions (see Figure 36.27) or double minute chromosomes (see Figure 36.28). In some cases, the amplified region contains a known oncogene or a gene related to one. In other cases, where amplification is not visible, the use of batteries of probes representing oncogenes shows that a particular oncogene is amplified. Examples of oncogenes that are amplified in various tumors include *c-myc, c-abl, c-myb, c-erbB, and c-K-ras.*

Established cell lines are prone to amplify genes (it is one of several karyotypic changes to which they are susceptible). All the same, the presence of known oncogenes in the amplified regions, and the consistent amplification of particular oncogenes in many independent tumors of the same type, again strengthens the correlation between increased expression and tumor growth. Of course, it is possible that the gene amplification gives an advantage to growth of the established tumor; it is not necessarily an event involved in its initiation.

Some proto-oncogenes are activated by events that change their expression, but which leave their coding sequence unaltered. The best characterized is *c-myc*, whose expression is elevated by several mechanisms. One common mechanism is the insertion of a nondefective retrovirus in the vicinity of the gene.

The ability of a retrovirus to transform without expressing a *v-onc* sequence was first noted during analysis of the bursal lymphomas caused by the transformation of B lymphocytes with avian virus. Similar events occur in the induction of T cell lymphomas by murine leukemia virus. In each case, the transforming potential of the retrovirus seems to lie with its LTR rather than with a coding sequence.

In many independent tumors, the virus has integrated into the cellular genome within or close to the *c-myc* gene. The gene consists of three exons; the first represents a long nontranslated leader, and the second two code for the c-Myc

protein. Figure 39.8 summarizes the types of insertion at this locus. The retrovirus may be inserted at a variety of locations relative to the *c-myc* gene.

The simplest insertions to explain are those that occur within the first intron. The LTR provides a promoter, and transcription reads through the two coding exons. Transcription of *c-myc* under viral control differs in two ways from its usual expression: the level of expression is increased (because the LTR provides an efficient promoter); and the transcript lacks its usual nontranslated leader (which may usually limit expression).

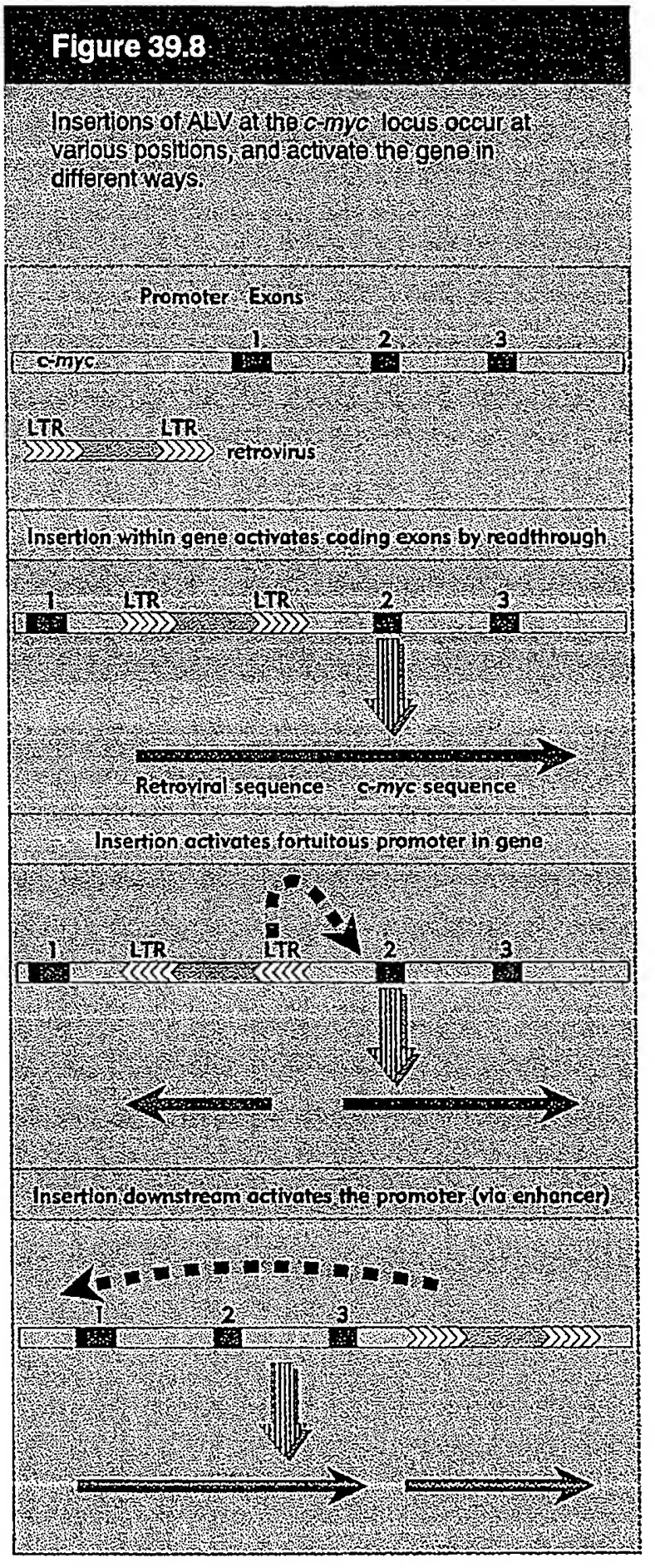
Activation of *c-myc* in the other two classes of insertions reflects different mechanisms. The retroviral genome may be inserted within or upstream of the first intron, but in reverse orientation, so that its promoter points in the wrong direction. Probably the LTR provides an enhancer that acts on an upstream sequence that fortuitously resembles a promoter. The retroviral genome also may be inserted downstream of the *c-myc* gene, in which case transcription initiates at the usual *c-myc* promoter(s), but is increased by the enhancer in the retroviral LTR.

In all of these cases, the coding sequence of c-myc *is unchanged, so oncogenicity is attributed to the loss of normal control, and increased expression, of the gene.*

Other oncogenes that are activated in tumors by the insertion of a retroviral genome include *c-erbB, c-myb, c-mos, c-H-ras,* and *c-raf.* Up to 10 other cellular genes (not previously identified as oncogenes by their presence in transforming viruses) are implicated as potential oncogenes by this criterion. The best characterized among this latter class are *wnt1* and *int2.* The *wnt1* gene codes for a protein involved in early embryogenesis that is related to the *wingless* gene of *Drosophila; int2* codes for a growth factor related to FGF.

Translocation to a new chromosomal location is another of the mechanisms by which oncogenes are activated. Certain chromosomal translocations are consistently associated with activation of oncogenes that lie near the breakpoints. This situation was originally discovered via a connection between the loci coding immunoglobulins and the

Figure 39.8



occurrence of certain tumors. Specific chromosomal translocations are often associated with plasmacytomas in the mouse and with Burkitt lymphomas in man. These tumors arise from undifferentiated B lymphocytes. The common feature in both species is that an oncogene on one chromosome is brought into the proximity of an Ig locus on another chromosome. Similar events occur in T lymphocytes to bring oncogenes into the proximity of a TcR locus.

The basic cause of the translocation event is a malfunction of the system responsible for recombining V and C genes or switching between IgH C genes. Instead of acting on two sites within the Ig or TcR locus, the system recombines the immune locus with an unrelated region on a different chromosome. This results in a **reciprocal translocation**, which is illustrated in **Figure 39.9**.

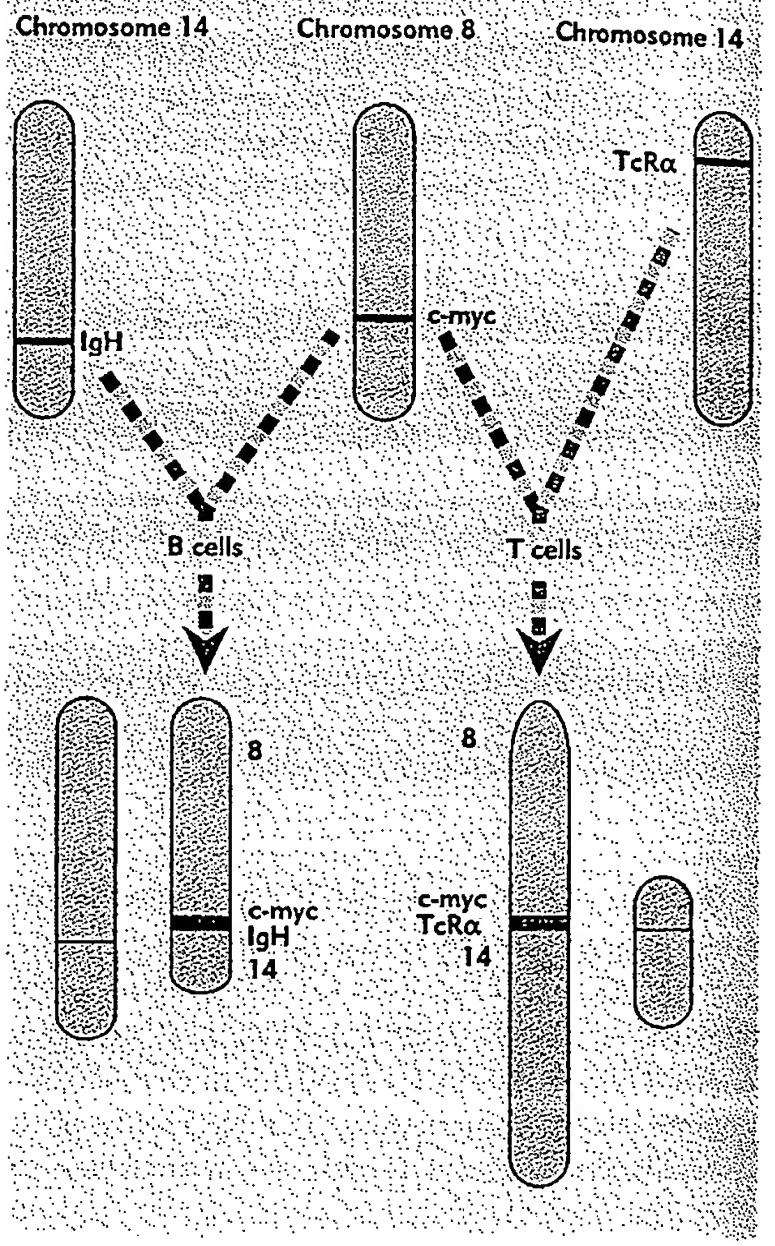
We do not know the basis for the involvement of the nonimmune partner, but in both man and mouse it is often the *c-myc* locus. In man, the translocations in B cell tumors usually involve chromosome 8, which carries *c-myc*, and chromosome 14, which carries the IgH locus; ~10% involve chromosome 8 and either chromosome 2 (κ locus) or chromosome 22 (lambda locus). The translocations in T cell tumors often involve chromosome 8, and either chromosome 14 (which has the TcRα locus at the other end from the Ig locus) or chromosome 7 (which carries TcR β locus). Analogous translocations occur in the mouse.

Translocations at the IgH locus in B cells fall into two classes. One type is similar to those observed at other Ig loci and at TcR loci, involving the consensus sequences used for V-D-J somatic recombination of active Ig genes. In the other type, the translocation occurs at a switching site, so these cases may be associated with function of the system that switches expression from one C_H gene to another.

When *c-myc* is translocated to the Ig locus, its level of expression is usually increased. The increase varies considerably among individual tumors, generally being in the range from 2 to 10. Why does translocation activate the *c-myc* gene? The translocation event does not involve fixed sites,

Figure 39.9

A chromosomal translocation is a reciprocal event that exchanges parts of two chromosomes. Translocations that activate the human *c-myc* proto-oncogene involve Ig loci in B cells and TcR loci in T cells.



but occurs at a variety of locations within a general region on each recombining chromosome. The event has two consequences: *c-myc* is brought into a new region, one in which an Ig or TcR gene was actively expressed; and the structure of the *c-myc* gene may itself be changed (but usually not involving the coding regions). It seems likely that several different mechanisms can activate the *c-myc* gene

in its new location (just as retroviral insertions activate *c-myc* in a variety of ways).

The correlation between the tumorigenic phenotype and the activation of *c-myc* by either insertion or translocation suggests that continued high expression of the c-Myc protein is oncogenic. Expression of *c-myc* must be switched off to enable immature lymphocytes to differentiate into mature B and T cells; failure to turn off *c-myc* maintains the cells in the undifferentiated (dividing) state. The oncogenic potential of *c-myc* has been demonstrated directly by the creation of transgenic mice carrying a normal *c-myc* gene linked to an enhancer. Transgenic mice carrying a *c-myc* gene linked to a B lymphocyte-specific enhancer (the IgH enhancer) develop lymphomas. The tumors represent both immature and mature B lymphocytes, suggesting that over-expression of *c-myc* is tumorigenic throughout the B cell lineage. Transgenic mice carrying a *c-myc* gene under the control of the LTR from a mouse mammary tumor virus, however, develop a variety of cancers, including mammary carcinomas. This suggests that increased or continued expression of *c-myc* transforms the type of cell in which it occurs into a corresponding tumor. Specificity of the tumor type may therefore depend on the mechanism used to activate *c-myc*; it is not an intrinsic property of the gene.

c-myc exhibits three means of oncogene activation: retroviral insertion, chromosomal translocation, and gene amplification. The common thread among them is increased expression of the oncogene rather than a qualitative change in its coding function, although in at least some cases the transcript has lost the usual (and possibly regulatory) nontranslated leader. *c-myc* provides the paradigm for oncogenes that may be effectively activated by increased (or possibly altered) expression.

Every translocation generates reciprocal products; sometimes a known oncogene is activated in one of the products, but in other cases it is not evident which of the reciprocal products has responsibility for oncogenicity. Also, it is not axiomatic that the gene(s) at the breakpoint have responsibility; for example, the translocation could provide an enhancer that activates another gene nearby.

A variety of translocations found in B and T cells have identified new oncogenes. In some cases, the translocation generates a hybrid gene, in which an active transcription unit is broken by the translocation. This has the result that the exons of one gene may be connected to another. In such cases, there are two potential causes of oncogenicity: the proto-oncogene part of the protein may be activated in some way that is independent of the other part, for example, because it is over-expressed under its new management (a situation directly comparable with the example of *c-myc*); or the other partner in the hybrid gene may have some positive effect that generates a gain-of-function in the part of the protein coded by the proto-oncogene.

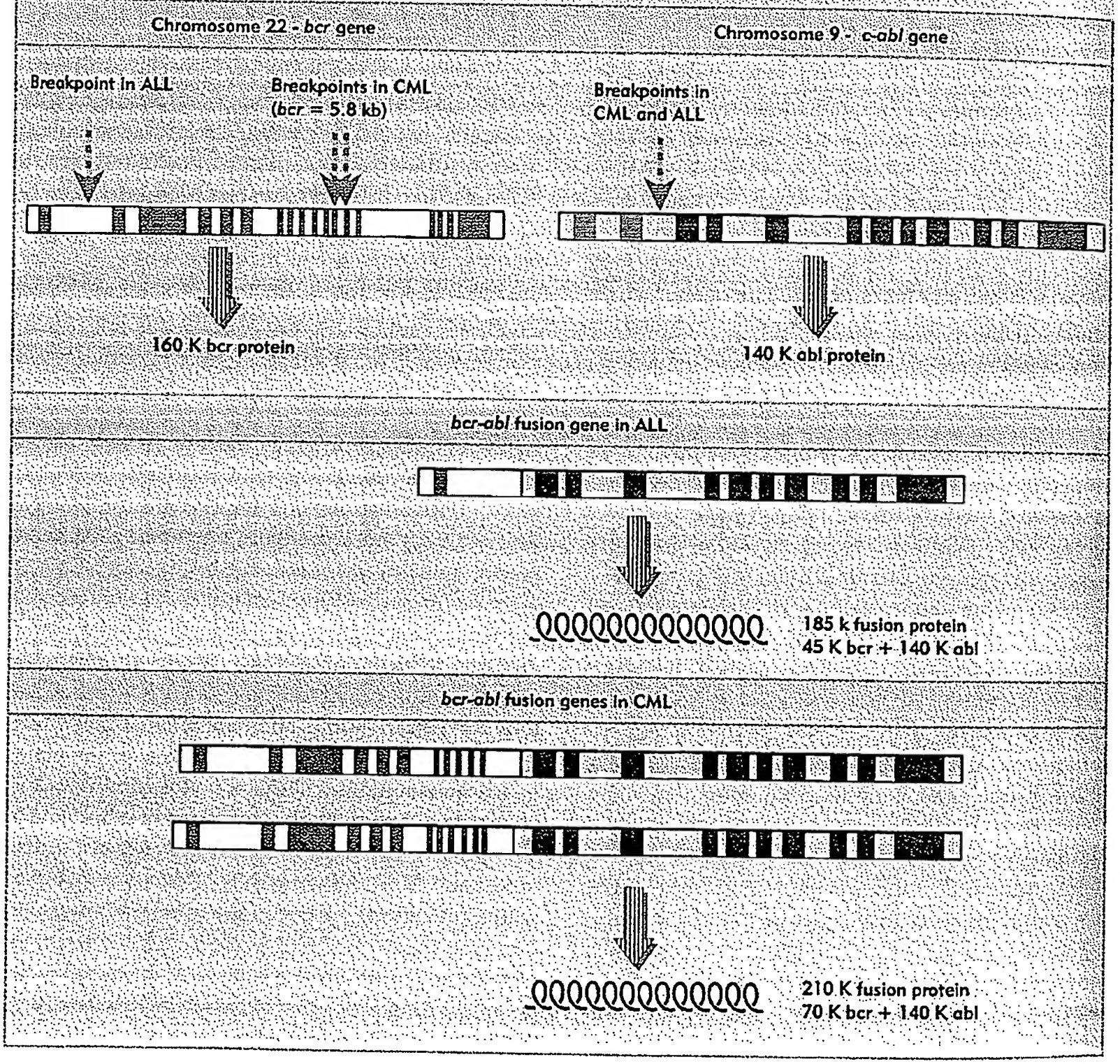
One of the best characterized cases in which a translocation creates a hybrid oncogene is provided by the *Philadelphia (PH¹)* chromosome present in patients with chronic myelogenous leukemia (CML). This reciprocal translocation is too small to be visible in the karyotype, but links a 5,000 kb region from the end of chromosome 9 carrying *c-abl* to the *bcr* region of chromosome 22. The *bcr* (*b*reakpoint *c*luster *r*egion) was originally named to describe a region of ~5.8 kb within which breakpoints occur on chromosome 22. Different cases of CML have breakpoints at different locations within this region.

The consequences of this translocation are summarized in **Figure 39.10**. The *bcr* region lies within a large (>90 kb) gene, which is now known as the *bcr* gene. The breakpoints in CML usually occur within one of two introns in the middle of the gene. The same gene is also involved in translocations that generate another disease, ALL (acute lymphoblastic leukemia); in this case, the breakpoint in the *bcr* gene occurs in the first intron.

The *c-abl* gene is expressed by alternative splicing that uses either of the first two exons. The breakpoints in both CML and ALL occur in the intron that precedes the first common exon. Although the exact breakpoints on both chromosomes 9 and 22 vary in individual cases, the

Figure 39.10

Translocations between chromosome 22 and chromosome 9 generate Philadelphia chromosomes that synthesize *bcr-abl* fusion transcripts that are responsible for two types of leukemia.



common outcome is the production of a transcript coding for a Bcr–Abl fusion protein, in which N-terminal sequences derived from *bcr* are linked to *c-abl* sequences. In ALL, the 185,000 dalton fusion protein has ~45,000 daltons of the Bcr protein linked to c-Abl; in CML the fusion protein of 210,000 daltons has ~70,000 daltons of the Bcr protein. In each case, the fusion protein contains ~140,000

daltons of the usual ~145,000 c-Abl protein, that is, it has lost just a few N-terminal amino acids of the c-Abl sequence.

Why is the fusion protein oncogenic? It relies on an interaction between the N-terminal region provided by *bcr* with the c-Abl protein. The *bcr* gene has a variety of sequence motifs related to proteins involved in signalling pathways, but the pertinent one is a serine/threonine kinase activity that is coded within the first exon. This autophosphorylates residues in this part of the protein, and the phosphorylation confers the ability to interact with a region of the c-Abl protein called the SH2 domain (we discuss the nature of SH2 domains later). This raises the possibility that the Bcr part of the protein interacts with the c-Abl sequences, perhaps changing their conformation and activating a latent oncogenic potential.

Changes at the N-terminus are involved in the activation of oncogenic activity of *v-abl*, a transforming version of the gene carried in a retrovirus. The *c-abl* gene codes for a tyrosine kinase activity; this activity is essential for transforming potential in oncogenic variants. Deletion (or replacement) of the N-terminal region activates the kinase activity and transforming capacity. So the N-terminus provides a domain that usually regulates kinase activity; its loss may cause inappropriate activation.

Loss of tumor suppressors causes tumor formation

The common theme in the role of oncogenes in tumorigenesis is that increased or altered activity of the gene product is oncogenic. Whether the oncogene is introduced by a virus or results from a mutation in the genome, it is dominant over its allelic proto-oncogene(s). A mutation that activates a single allele is tumorigenic. Tumorigenesis then results from gain of a function.

Certain tumors are caused by a different mechanism: loss of both alleles at a locus is tumorigenic. Propensity to form such tumors may be inherited through the germline; it also occurs as the result of somatic change in the individual. Tumorigenesis then results from loss of function. Such cases identify tumor suppressors: genes whose products are needed for normal cell function, and whose loss of function causes tumors. The two best characterized genes of this class code for the proteins RB and p53.

Retinoblastoma is a human childhood disease, involving a tumor on the retina. It occurs both as a heritable trait and sporadically (by somatic mutation). It is often associated with deletions of band q14 of human chromosome 13. The *RB* gene has been localized to this region by molecular cloning.

Figure 39.11 illustrates the situation. Retinoblastoma arises when both copies of the *RB* gene are inactivated. In the inherited form of the disease, one parental chromosome carries an alteration in this region, usually a deletion. A somatic event in retinal cells that causes loss of the other copy of the *RB* gene causes a tumor. In the sporadic form of the disease, the parental chromosomes are normal, and both *RB* alleles are lost by (individual) somatic events.

Almost half the cases of retinoblastoma show deletions at the *RB* locus. In other cases, transcripts of the locus are either absent or altered in length. The protein product is absent from retinoblastoma cells. The cause of the tumor is therefore loss of protein function, usually resulting from mutations that prevent gene expression (as opposed to point mutations that affect function of the protein product). Loss of *RB* is involved also in other forms of cancer, including osteosarcomas and small cell lung cancers.

What is the molecular function of RB? It interacts with a variety of other proteins, including several tumor antigens: SV40 T antigen, adenovirus E1A, human papilloma virus E7. One possibility is that part of the oncogenicity of these proteins is due to

REVIEW ARTICLE Medical Biology 62: 304–317, 1984

AMPLIFICATION OF CELLULAR ONCOGENES IN CANCER CELLS

K. ALITALO

FROM THE DEPARTMENT OF VIROLOGY, UNIVERSITY OF HELSINKI, HELSINKI, FINLAND

ABSTRACT

Regulatory or structural alterations of cellular oncogenes have been implicated in the causation of various cancers. Oncogene alteration by point mutations can result in a protein product with strongly enhanced oncogenic potential. Aberrant expression of cellular oncogenes may be due to tumour-specific chromosomal translocations that dysregulate the normal functions of a proto-oncogene. Amplification of cellular oncogenes can also augment their expression by increasing the amount of DNA template available for the production of mRNA. It appears that amplification of certain oncogenes is a common correlate of the progression of some tumours and also occurs as a rare sporadic event affecting various oncogenes in different types of cancer. Amplified copies of oncogenes may or may not be associated with chromosomal abnormalities signifying DNA amplification: double minute chromosomes and homogeneously staining chromosomal regions. Amplified oncogenes, whether sporadic or tumour type-specific, are expressed at elevated levels, in some cases in cells where their diploid forms are normally silent. Increased dosage of an amplified oncogene may contribute to the multistep progression of at least some cancers.

KEY WORDS: CELLULAR ONCOGENES, GENE AMPLIFICATION, MULTISTEP CARCINOGENESIS, CLONAL SELECTION, KARYOTYPIC ABNORMALITIES, DOUBLE MINUTE CHROMOSOMES, HOMOGENEOUSLY STAINING CHROMOSOMAL REGIONS

DNA SEQUENCE AMPLIFICATION AND CYTOGENETIC ABNORMALITIES IN TUMOURS

Since its discovery in drug-resistant eukaryotic cells, somatic amplification of specific genes has been implicated in an increasing variety of adaptive responses of cells to environmental stresses (70, 79). Cytogenetic abnormalities, double minute chromosomes (dmin:s) associated with DNA amplification had already been discovered in tumour cells before the discovery of dmin:s and homogeneously staining chromosomal regions (HSR:s) in cells selected for drug-resistance (12, 24, 49, 50, 56). In metaphase spreads, dmin:s appear as small, spherical, usually paired chromosome — like structures that lack a centromere (Fig. 2). HSR:s stain with intermediate intensity throughout their length rather than with the normal pattern of alternating dark and light bands in both trypsin-Giemsa (Fig. 3A) and quinacrine dihydrochloride-stained preparations. Both kinds of abnormalities are occasionally found in metaphases of freshly isolated cancer cells but not of normal cells (8).

Dmin:s and HSR:s are apparently rare in tumour cells in vivo, although exact data are difficult to obtain since the abnormalities are easily missed in routine cytogenetic analysis (8, 42). Dmin:s and HSR:s have been described in most types of in vitro-cultured malignant tumour cells, with a notable frequency in neuroblastoma cell lines (11). Initial growth in cell culture apparently selects for tumour cells that contain either dmin:s or HSR:s. Moreover, in culture dmin:s are frequently lost, concomitant with the appearance of clonal populations of cells that have developed an HSR, suggesting that the two cytogenetic abnormaltities are alternative forms of gene amplification and that HSR:s may confer a selective advantage on cells over dmin:s (11, 70). It has been assumed that HSR:s can break down to firm dmin:s and that dmin:s can integrate into chromosomes to generate HSR:s (11, 23). Amplified genes may also occupy abnormally banding regions, ABR:s (51, 59). Experimental work on drug-resistant cells has shown that in the absence of a selection pressure (drug), dmin:s and the amplified genes that they contain are lost, whereas amplified DNA in the form of HSR:s is retained in the cells (71). This is explained by the fact that dmin:s are segregated unevenly in mitosis and frequently get lost from the nucleus due to

TABLE 1

Currently known oncogenes.

Retrovirus (example)	Oncogene	Gene product: Cellular location	Gene product: Function of protein	Class
ONCOGENES FOUND IN RETROVIRUSES				
RSV	*src*	Plasma membrane	Tyrosine-specific protein kinases (*fgr* contains sequences homologous to actin)	Class Ia (Cytoplasmic tyrosine protein kinases)
Y73V	*yes*	Plasma membrane		
GR-FeSV	*fgr*	Plasma membrane		
Ab-MuLV	*abl*	Plasma membrane		
FuSV	*fps/fes*	Cytoplasm (plasma membrane?)		
ST-and GA-FeSV	*fes/fps*	Cytoplasm (cytoskeleton?)		
UR2V	*ros*			
AEV	*erb*-B	Plasma membrane and cytoplasmic membranes	EGF receptor's cytoplasmic domain	Class 1b (Class 1a-related proteins)
SM-FeSV	*fms*	Plasma membrane and cytoplasmic membranes	Cytoplasmic domain of a growth factor receptor?	
MH-2V	*mil/raf*	Cytoplasm	?	
3911-MSV	*raf/mil*	Cytoplasm	?	
Mo-MSV	*mos*	Cytoplasm	?	
SSV	*sis*	Secreted	PDGF-like growth factor	Class 2 (Growth factors)
Ha-MSV	Ha-*ras*	Plasma membrane	GTP-binding proteins	Class 3 (Cytoplasmic GTP:ases)
Ki-MSV	Ki-*ras*	Plasma membrane		
FBJ-MuSV	*fos*	Nucleus	?	Class 4 (Nuclear phosphoproteins)
OK-10V	*myc*	Nucleus	Nuclear matrix protein	
AMV	*myb*	Nucleus	?	
SKV 770	*ski*	Nucleus?	?	Unclassified
REV	*rel*	?	?	
AEV	*erb*-A	?	?	
E26V	*ets*	?	?	
ONCOGENES FOUND IN TUMOUR CELLS BUT NOT IN RETROVIRUSES				
Tumour cell				
Neuroblastoma	N-*ras*	Plasma membrane	GTP—binding	Class 3
Neuroblastoma	N-*myc*	?	?	Class 4
Small-cell lung cancer	L-*myc*	?	?	Class 4
Neuro-/Glioblastomas	*neu*	Plasma membrane	Growth factor receptor	Class 1b

their lack of centromeres, (49). HSR chromosomes carry centromeres and are therefore divided equally at mitosis. If dmin:s and HSR:s contain amplified genes that encode growth-stimulating protein products, it would follow that the more stable chromosomal form, the HSR, confers a greater selective growth advantage for cells. Although dmin:s and HSR:s have been described predominantly in tumour cells selected for resistance to cytotoxic drugs, it is also clear that dmin:s and HSR:s may be present in cancer cells before any form of therapy (8). It was in this setting that we and others first chose to explore the possible amplification of cellular oncogenes. (By definition, cellular oncogenes are normally innocent genetic loci which can be activated to transforming genes in various ways).[1]

DMIN:S AND HSR:S CONTAIN AMPLIFIED ONCOGENES

Table 2 summarizes the somatic amplifications of cellular oncogenes so far reported in

[1] It is not the purpose of this review to deal with all forms of DNA damage that have been found to activate cellular oncogenes. For the purpose of integrating the review into a coherent picture, however, the reader is given a list of known cellular oncogenes in Table 1 and the schematic Figure 1 illustrating the various ways in which the oncogenic potential of different proto-oncogenes can be activated.

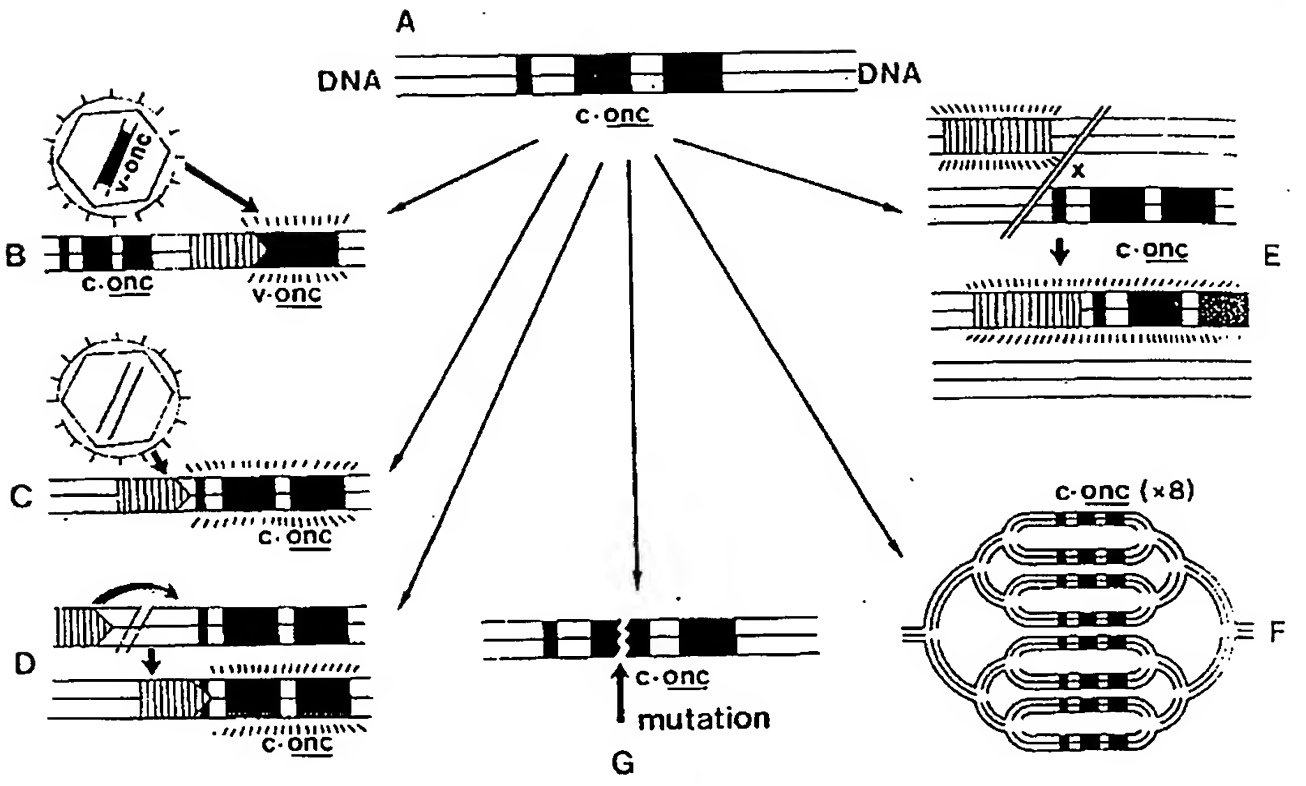


Fig. 1. Activation of cellular oncogenes. The haploid complement of a proto-oncogene is schematically depicted in A, composed of three exons (black boxes) in a segment of DNA. The different activated forms are schematically outlined in B—G. The abbreviation c-*onc* stands for cellular oncogene, v-*onc* viral oncogene, DNA sequences with associated strong promoter/enhancer functions are striated, and an actively transcribed gene is marked with radiations. B. Acute transforming retroviruses have the capacity to transduce cellular oncogenes (c-*onc*) into their genome, modify them and reinsert their activated oncogenes (v-*onc*) into the genome of host animal cells as a part of their proviral forms. The activity of the v-*onc* gene is greatly enhanced due to the associated promoter of the proviral long terminal repeat. Both increased dosage of the oncogene and structural mutations within its sequence may contribute to tumorigenesis. C. Slow transforming retroviruses without oncogenes replicate and reinsert their proviral copies into the host cell DNA during a latency period from infection to tumorigenesis. Tumor initiation through hyperplastic growth may begin, when the provirus integrates sufficiently close to a proto-oncogene to activate it through promoter or enchancer functions. It should be noted, however, that mutations have also been found in the oncogenes thus activated and that mutational damage to other oncogenes has been described in the resulting tumors. D. In some mouse plasmacytomas, a retrovirus-like DNA element (directing the synthesis of the so-called intracisternal A-type particles, IAPs) has been found in association with a transcriptionally activated oncogene c-*mos*. The IAP insertion also disrupts the 5' part of c-*mos* (64). E. In humans, as well as in animals, chromosome translocations may place proto-oncogenes into transcriptionally active regions of chromatin, where they may be activated. The details of this mechanism have not been worked out, but it is believed to occur for c-*myc* and c-*abl* genes in Burkitt lymphomas and Philadelphia-chromosome positive leukemias, respectively (35, 40). F. Increased amounts of oncogene-specific RNA and protein can also result from an excess of DNA template for transcription acquired through oncogene amplification. The present review concentrates primarily on this mechanism. G. Mutationally activated oncogenes have been found in nearly one fifth of human malignant tumours. Oncogene loci activated by somatic structural mutations are revealed by transfection experiments, where they are introduced into genetic background of nontumourigenic cultured immortalized cells. Several such transforming loci have been cloned and many of them belong to the c-*ras* oncogene family. It should be pointed out that both structural mutations and either increased expression or activation of a complementing oncogene may be required to achieve a fully tumorigenic phenotype (44).

tumour cells. Although the sampling of tumours is at present small, the finding of known cellular oncogenes among amplified DNA represented by dmin:s and HSR:s of cancer cells is provocative. Amplification has been found to affect at least five out of twenty known cellular oncogenes and the degree of gene amplification varies from five to many hundred-fold over the single haploid copies found in normal cells (see also ref. 18). The first amplification reported involved the c-*myc* oncogene (see Table 1) in a promyelocytic leukaemia cell line HL-60 (20, 25). The degree of c-*myc* amplification is between 8—32 fold both in the HL-60 cell line and in primary leukaemic cells from the patient (20, 25). Original clonal lines of HL-60 were later found to contain some dmin:s in culture but their number was insufficient to establish any clear correlation with amplified c-*myc*. Such a correlation, however, was discovered for c-*myc* amplification in a neuroendocrine cell line from

Fig. 2. Double minute chromosomes (arrowheads) in COLO 320DM colon carcinoma cells. A. The dmin:s are resolved as paired dots among normal chromosomes in this fluorescent, benzimidazole-stained preparation B—D. Purification of dmin:s by differential centrifugations. B. The starting material, C. Chromosome fraction, D. Purified dmin:s (Donna George and the author, unpublished data and ref. 52).

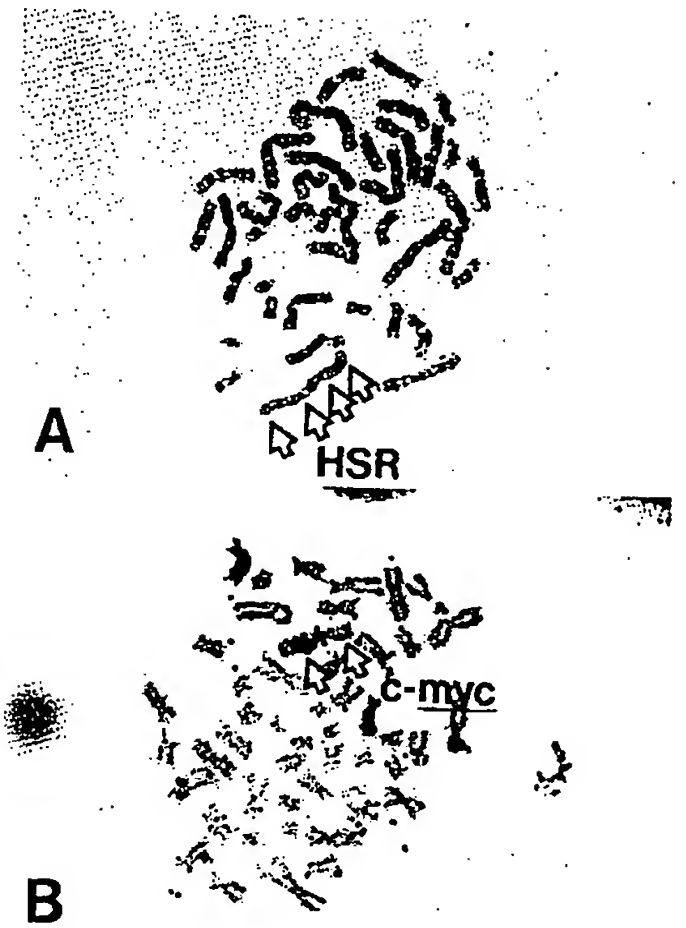


Fig. 3. A. The homogeneously staining regions (HSR) in the G-banded HSR-marker chromosome comprise a major portion of both its long and short arms. The HSR-marker chromosome has evolved from an X-chromosome (52 and unpublished data of C. C. Lin and the author). B. The about 30-fold amplified copies of the c-*myc* oncogene in COLO 320 cells were found to be located to dmin:s and HSR:s. The latter is shown here by in situ-hybridization (5, 52).

a colon carcinoma, COLO 320 (5). In these cells, the approximately 30-fold amplified c-*myc* copies were mapped either to HSR:s of a marker chromosome (5, Fig 3B) or to dmin:s (52), depending on the particular subline studied. Since dmin:s were already present in the primary tumour cells from this colon carcinoma (63), it is likely that c-*myc* had also been amplified during growth of the tumour in vivo. Similarly, amplified copies of the c-Ki-*ras* oncogene were mapped to dmin:s and HSR:s of a mouse adrenocortical tumor Y1 (74). An extensive search for changes in other oncogenes and tumour cells has since revealed amplifications that do not show up as dmin:s or HSR:s.

Thus, for example, the c-*myb* oncogene is amplified in a characteristic marker chromosome of a colon carcinoma without evidence of HSR:s (ref. 6, Fig. 4) and in other tumours, the amplified c-*abl* and c-*myc* oncogene loci map to abnormally banding regions (ABR:s) in translocated or resident chromosomal segments, respectively (59, 76).

TRANSLOCATIONS AND REARRANGEMENTS MAY ACCOMPANY ONCOGENE AMPLIFICATION

The evolution and progression of the karyotype of tumour cells is complicated (see ref. 68). Concomitant with amplification, DNA sequences acquire an increased mobility in the genome with extrachromosomal intermediates

TABLE 2

Sporadic and tumour-specific amplification of cellular oncogenes.[*]

Tumour cells	Oncogene	Fold amplified	Chromosomal location of amplified gene	Expression elevated	Remarks	References
Sporadic:						
HL60 (acute promyelocytic leukaemia, M3)	c-*myc*	20×	8q(ABR)	Yes	Amplification present in primary leukaemic cells	20, 25, 59
COLO320 (colon carcinoma)	c-*myc*	30×	dmin, HSR	Yes	Part of the amplified c-*myc* sequences rearranged	4, 5, 52
Y1 (adrenocortical tumour)	c-Ki-*ras*	50×	dmin, HSR	Yes	Levels of p $21^{c\text{-}Ki\text{-}ras}$ protein elevated	74
COLO201/205 (colon carcinoma)	c-*myb*	10×	mar1	Yes	Patient treated with 5-fluorouracil prior to culturing of the tumour cells	4, 6,88
K562 (chronic myelogenous leukaemia, CML)	c-*abl*	10×	mar(ABR)	Yes	C_h coamplified in the marker that may be derived from chromosome 22, c-*abl* protein-associated tyrosine kinase activated	21, 22, 41, 54, 76
A431 (epidermoid carcinoma)	c-*erb*B	15—20×	n.d.	Yes	Amplification linked to chromosome 7 translocation and sequence rearrangements Amount of protein product, the EGF receptor, elevated	82 (see 36)
ML1—3 (acute myeloid leukaemia, M2)	c-*myb*	5—10×	n.d.	Yes	Abnormalities of chromosome 6q22—24, where c-*myb* is normally located	34, 61, 91
SK BR-3 (breast carcinoma)	c-*myc*	10×	n.d.	Yes		43
SEWA (polyoma virus-induced mouse tumour)	c-*myc*	30×	n.d.	Yes	Cells have dmin:s depending on culture conditions; c-*myc* amplification correlates with growth as a tumour	Manfred Schwab, personal communication
Lu-65 (lung giant cell carcinoma)	c-*myc* c-Ki-*ras*	8× 10×	n.d. n.d.	n.d. n.d.	At least some copies of c-Ki-*ras* mutated	80
Primary leukemic cells from an acute myeloid leukemia (M2) patient	c-*myc*	33×	n.d.	n.d.		Unpublished data of the author and A. de la Chapelle
Tumour-specific:						
small-cell lung cancer	c-*myc* L-*myc*, N-*myc*	up to 80×	n.d.	Yes	Most amplifications in the variant phenotype of SCLC	53, 69
Neuroblastomas	*N-myc*	up to 250×	dmin, HSR	Yes	N-*myc* also amplified in primary tumours of advanced grade	14, 48, 72, 73,. 75
Glioblastomas	c-*erb*B	—	—	—	Rearrangements also found personal communication	Josef Schlessinger, personal communication

n.d. = not determined, mar = marker chromosome, M2, M3 refer to the French-American-British classification of acute myeloid leukemias.

[*] At least one case of oncogene amplification in normal germ-line cells has been found (18).

visualized as dmin:s, transpositions and translocations to other chromosomal segments, etc. (see 70 for references). There may not be preferred chromosomal sites for the apparent reintegration of dmin:s as HSR:s (75). In at least one case, however, an oncogene may have been caught amplifying in situ in its resident chromosomal site (59). The finding of moderately amplified oncogenes also in chromosomal sites lacking HSR:s suggests that

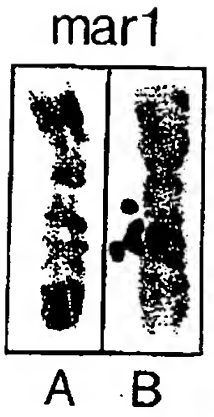


Fig. 4. Localization of amplified c-*myb* in COLO 201/205 cells by in situ hybridization. Shown is a characteristic, large marker chromosome (mar1) with G-banding (A) and associated c-*myb* autoradiographic grains (B). Note the absence of HSR:s. Marl has probably evolved from chromosome number 6, the resident site of the c-*myb* oncogene in normal cells (34, 88, 91). (Robert Winqvist and the author, unpublished data).

Fig. 5. Amplification and rearrangement of c-*myc* in COLO 320 cells. 10 µg of cellular DNA was digested with Sst I, electrophoresed, blotted and probed with a v-*myc* Pst I fragment (ref. 2, left panel). Fragments of 2.7 kbp and 1.4 kbp are seen in both normal and amplified c-*myc* DNA. The 3.3 kbp fragment is derived from a DNA segment of unknown origin translocated to the 5' region of c-*myc* with a concomitant deletion of its first exon (unpublished data of Manfred Schwab and the author). HSF, human skin fibroblasts; DM, COLO 320 DM cells; HSR, COLO 320 HSR cells. Different amounts of DNA from COLO 320 DM cells as indicated were mixed with calf thymus DNA to give 24 µg of total DNA, cleaved with Sst I, electrophoresed, blotted and probed with a fragment of 3'human c-*myc* sequences. The intensities of the 2.7 kbp c-*myc* fragment in different samples were compared to assess its copy number, estimated to be about 30 (5).

(onco)gene amplification may be more common than the structural alterations shown by chromosome banding and microscopy (6, 88).

In at least three cases reported amplification has been accompanied by a DNA rearrangement of the oncogene (5, 20, 82). In the colon carcinoma COLO 320 both damaged and normal versions of the c-*myc* gene are amplified (5). Although individual cell clones have not yet been examined, our unpublished experiments suggest that the same dmin-containing cells harbor and express both normal and rearranged forms of c-*myc*. The normal version of the amplified gene, however, predominates in COLO 320 cells containing HSR:s; the rearranged version is present only in what appears to be a single copy (Fig. 5). In the chronic myeloid leukaemia (erythroleukaemia) cell line K562 an amplified DNA segment consists of portions of both the c-*abl* oncogene and the immunoglobulin C_λ locus (76). In both cases abnormal transcripts are produced from the rearranged amplified oncogenes (Fig. 6 and ref. 22). In K562 cells, the abnormal c-*abl* oncogene product has also been activated as a tyrosine protein kinase (41). It is not known, however, whether structural alterations of the genes preceded amplification or whether they were acquired during the process of gene amplification. It seems likely that a chromosomal translocation of c-*abl* to the C_λ locus preceded DNA amplification in the K562 cells, since all amplified copies were also rearranged (21), with the change reminiscent of the Philadelphia translocation (t(9, 22)) found in most chronic myeloid leukaemia tumours (35, 66—68). Although they have not been sequenced, other reported cases of amplified oncogenes are apparently normal on basis of mapping with restriction endonucleases (see Table 2). Therefore we cannot at present view mutation as a necessary companion of oncogene amplification.

THE MECHANISMS OF GENE AMPLIFICATION

The mechanisms of gene amplification and the structure of the amplified DNA have been worked out mainly in experimental settings involving selection for drug-resistance in cell culture (70). Although the mechanisms are still incompletely known and may vary in different cases, some general features have emerged.

A spontaneous degree of illegitimate DNA replication seems to exist in normal cells so that various segments of DNA are replicated more than once during a single cell cycle (37). In unselective conditions this DNA is probably lost e.g., through formation of micronucleae

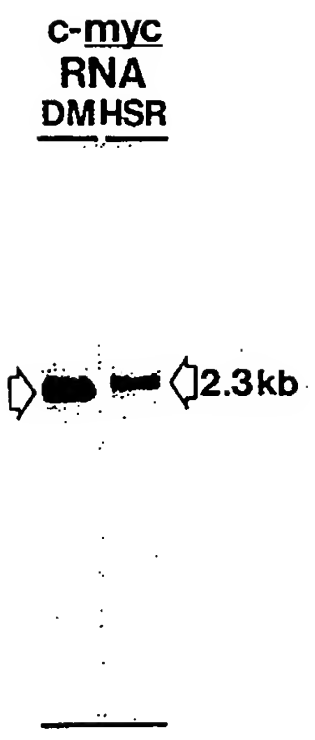


Fig. 6. Comparison of the electrophoretic mobilities of c-*myc* mRNA:s from COLO 320 DM and HSR cells. The size of the normal c-*myc* mRNA is 2.3 kb. The rearranged c-*myc* locus in DM cells (see Fig. 5) seems to be predominantly expressed giving rise to a shortened RNA.

because the newly synthesised extra copies of DNA are not covalently linked to chromosomal DNA of mitotic cells (65, 71). If, however, there is a selection pressure to retain an increased gene dosage, progressive multiplication of gene copy number results. The incidence of cells bearing amplified genes under conditions of cytotoxic selection can vary by two orders of magnitude and is greatly increased by the presence of mitogenic substances (hormones or tumour promoters) during selection (10, 84, 85) or certain carcinogenic or cytotoxic agents before selection (15, 55, 79, 80, 81, 85). An interesting hypothesis suggested by Varshavsky (84, 85) supposes that the origins of DNA replication "fire" (initiate replication) illegitimately several times during a single cell cycle and that this kind of "replicon misfiring" may be increased by substances such as tumour promoters and mitogenic hormones (10, 84, 85). Mariani and Schimke (55) point out that most of the cytotoxic agents that increase the incidence of gene amplification are inhibitors of DNA synthesis. Aberrant replication is known to take place after transient inhibition of DNA synthesis and this response can lead to gene amplification (46, 47, 55, 90). Mitogenic hormones probably increase disproportionate DNA replication, but they also enhance the colony forming efficiency of drug-resistant cells in selective conditions (10).

According to the studies of Axel and his collaborators (65), the multiple cycles of unscheduled DNA replication at a single locus during a single cell cycle result in a structure schematically outlined in Fig. 1F. The hydrogen-bonded amplified copies of DNA depicted in Fig. 1F must resolve into a tandem linear array before the next mitosis. This may well occur by homologous recombination between any one of several repeated sequences within the amplified domain (45, 65). Part of the recombinations would lead to extrachromosomal circles possessing an origin for replication (16, 62); these could be the precursors of dmin:s. The unequal recombinations mean that the resolved linear structure consists of tandemly repeated but heterogeneous units. According to Axel's model a gradient of amplification is formed so that centrally located sequences are amplified more than sequences distal to the origin of replication (65). This also has, in fact, been found to explain the large, complex DNA domain amplified in neuroblastoma cells in vivo (38, see also below).

The chromosomal site of integration of transfected genes significantly affects the frequency and cytogenetic result of their experimentally induced amplification (83). The amplification frequency in some transformants has been found to be 100-fold that of the others (83). This suggests that there also are preferred chromosomal positions for amplification of host cellular genes and that chromosomal rearrangements may facilitate gene amplification by positioning chromosomal sequences in a favorable array. In respect of the structural properties of the sequences involved in gene amplification, recombinatorially active regions have been implicated in experimental cases. DNA rearrangements involving restriction fragment length polymorphisms and variation in gene copy number have been detected in the human genome between clusters of short repetitive interspersed DNA sequences called Alu family DNA-sequences (17). Such inter-Alu sequences have also been detected in an extrachromosomal DNA form, including covalently closed circles (17, 78). The copy number of inter-Alu sequences apparently varies in an age- and tissue-specific manner (17, 78), but any comprehensive analysis of the phenomenon in human tumours is not yet available. It is also not yet clear whether these kinds of repetitive sequences are involved in generating amplified oncogene sequences in dmin:s or HSR:s in tumours.

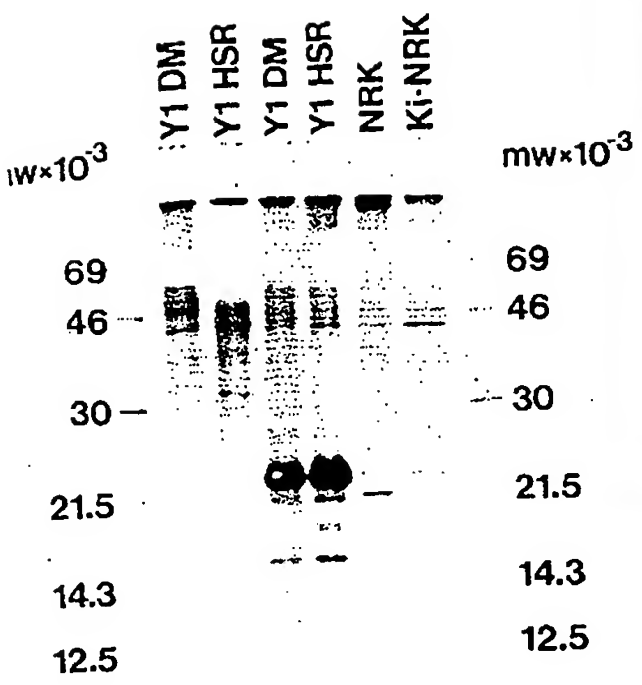


Fig. 7. Elevated levels of the $p21^{c\text{-}Ki\text{-}ras}$ protein in Y1 cells (74). The Y1 DM and HSR cells which harbor a 50-fold amplified c-Ki-*ras* oncogene (74) and control cells were labeled with [^{35}S]-methionine and the $p21^{c\text{-}Ki\text{-}ras}$ protein was immunoprecipitated, as detailed (74), with normal rat serum (NRS) or rat monoclonal anti-p21 serum. The proteins were electrophoresed in a 15 % polyacrylamide gel in the presence of SDS. In addition to a major p21 band, a labeled band at about 16 kd was present in the immunoprecipitates. The amount of radioactivity in p21 was about 50 fold that in normal rat kidney cells. The Kristen sarcoma virus-transformed rat kidney cells (obtained from the American Tissue Culture Collection) also yielded unexpectedly low amounts of the v-Ki-*ras* protein.

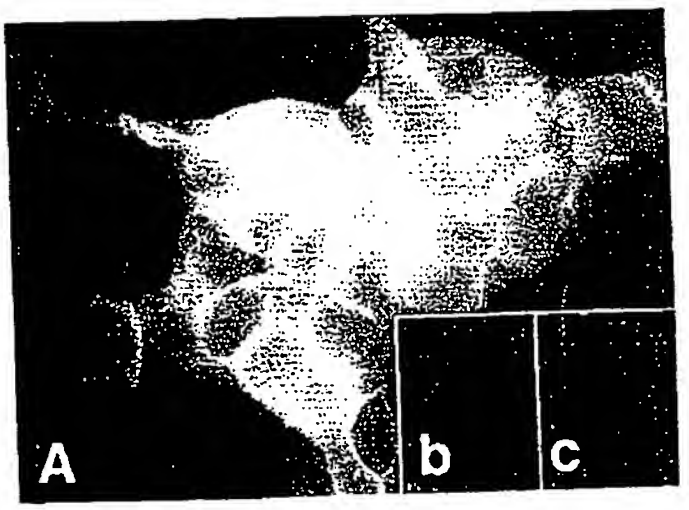


Fig. 8. A. Indirect immunofluorescence for $p21^{c\text{-}Ki\text{-}ras}$ in Y1 DM cells. Similar fluorescence of the plasma membrane was obtained for the Y1 HSR cells. Inset (b) shows control staining with normal rat serum and inset (c) staining of normal rat kidney cells with the monoclonal antibody against p21.

CARCINOGEN-INDUCED GENE AMPLIFICATION AND CLONAL SELECTION OF CANCER CELLS

Although cell sorting experiments have shown a basal spontaneous rate of gene amplification in eukaryotic cells (37), this can be increased severalfold by metabolic inhibitors or cytotoxic agents (15, 37, 70, 81, 85). In many respects the latter response is reminiscent of the so-called SOS-response elicited in bacteriae by noxious stimuli (see 28). In a teleological context, the rapid induction of gene amplification that apparently occurs frequently through extrachromosomal intermediates may provide cells with genetic material for subsequent selective pressures operating in harmful conditions (60). In cancer cells, the mechanism may enhance the emergence of clonal populations of cells with increasingly malignant properties (58, 60). Such genetic instability of cancer cells is clearly enhanced, leading to the rapid evolution of increasingly malignant tumour cell populations (19, 58). A serious question of practical importance is whether drug resistance in treated patients also selects cells that have an enhanced ability to amplify (onco)genes important for growth and progression of the tumour (84, 85). It is also possible that some of the carcinogenic insults caused by mutagens are only expressed as a result of subsequent amplification events induced by tumour promoters (84, 85) or facilitated by hormones in, say replicating epithelial cells (10). The persistence of dmin:s in some tumours suggests that there is a selection pressure for their retention (8, 9, 11, 23). Amplified DNA in dmin:s must contain an origin for DNA replication (62) and must be selected for in daughter cell populations, where it is unevenly segregated (71). In the absence of such a selection pressure dmin:s are lost (71). In at least one study the length of an HSR has been found to increase during a selection of malignant cells for enhanced tumourigenicity (30).

The amplified c-*erb*B gene in A431 cells codes for epidermal growth factor receptor (27). The abundant amounts of receptor protein on A431 cell surface may, however, provide the cells with an abnormal growth response (31). A naive supposition is that the amplified sequences in dmin:s and possibly in HSR:s of tumours contain growth-promoting genes (see 36 for references). This seems to fit well with

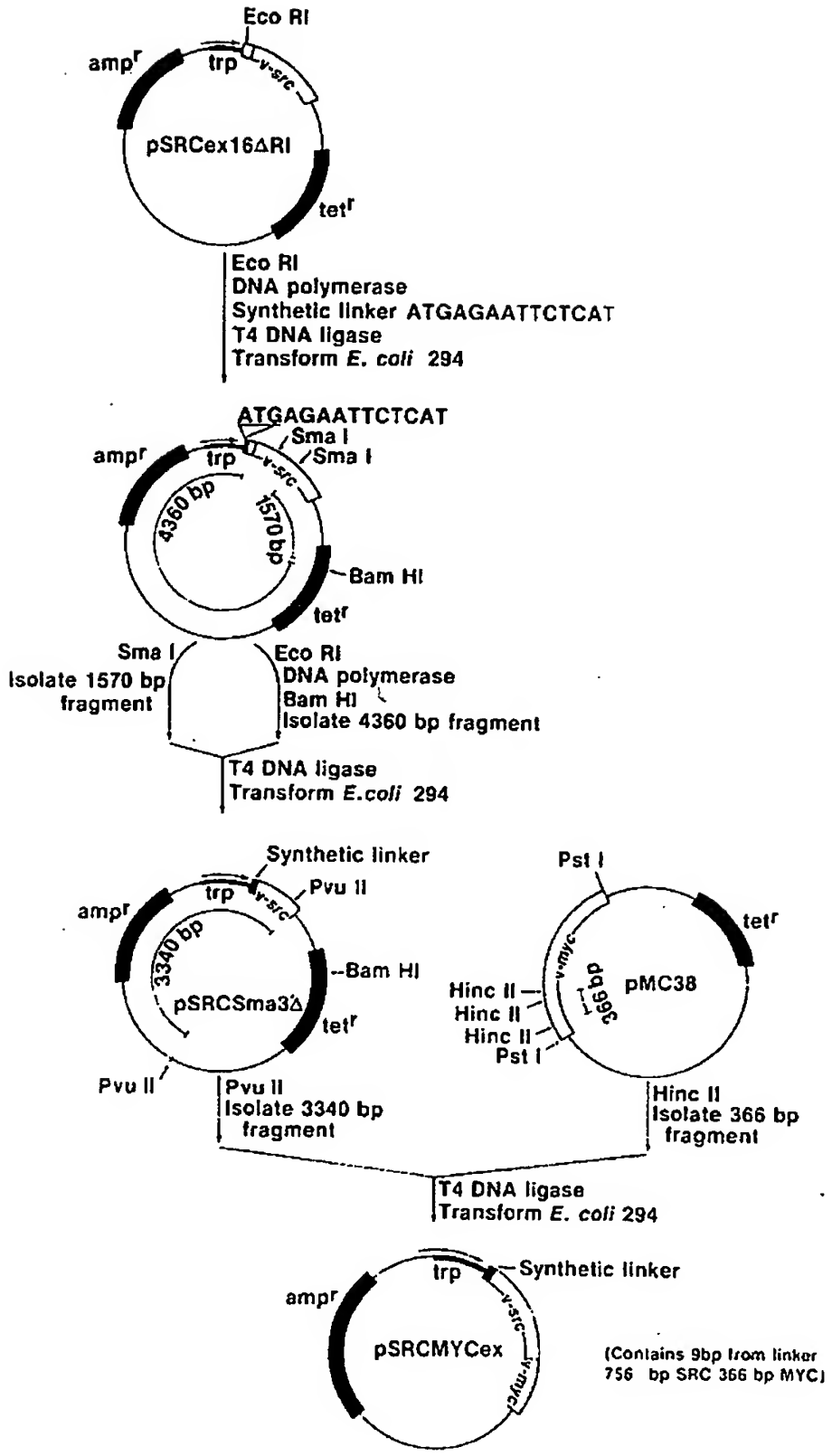


Fig. 9. Construction of a v-*myc* expression vector. A synthetic linker (ATGAGAATTCATCAT) containing a translational initiation codon was inserted downstream from the *trp* promoter in the pSRC ex16 RI expression vector described previously (see ref. 3). Approximately one half of the v-*src* sequences coding for the aminoterminal portion of pp60$^{v\text{-}src}$ protein were then deleted and the remaining portion ligated in translational codon frame with the synthetic ATG. A Hinc II fragment of v-*myc* from plasmid clone MC 38 (nucleotides 320—685 in the v-*myc* sequence in ref. 2) was ligated downstream form remaining v-*src* sequences in continuity with its reading frame. The resulting product contained 3 amino acids from the synthetic linker, 252 amino acids encoded by the 756 base pair fragment from Sma I to Pvu II restriction sites in v-*src* DNA, 122 amino acids from the v-*myc* and 6 amino acids (corresponding to nucleotides 2968—2085) from the pBR322 vector (3).

recent findings on amplified oncogenes, though in many cases the search for an amplified oncogene is still continuing. Even positive findings do not mandate a role for amplified cellular oncogenes, however, because the domain of amplified DNA is inevitably much larger than a single genetic locus (e.g. 38).

ENHANCED EXPRESSION OF AMPLIFIED ONCOGENES

In all cases where they have been studied, the amplified oncogenes have been found abundantly expressed at the RNA level, roughly in proportion to the amount of DNA amplification (see Table 1). Described cases of elevated RNA expression include examples of abnormal (5, 22) and ectopic (6) transcription. In at least four cases this enhancement is not limited to synthesis of RNA (31, 33, 41, 74, 82). The Y1 cells that have amplified c-Ki-*ras* contain exceptionally large amounts of its protein product situated on the plasma membrane (ref. 74, Fig. 7 and 8). High amounts of the c-*myc* encoded protein are also found in COLO 320 cells that have amplified the gene (33). The *myc* oncogenes have recently been shown to encode nuclear proteins (ref. 1, 3, 26, 29, 32, 33, Fig. 9—11). Both the expression of the c-*myc* mRNA (39) and the subcellular localization of myc proteins are linked to the cell cycle (ref. 89, Fig. 12). It may be that elevated expression of specific c-*myc* functions is necessary for cell cycle progression and the growth transformation aspect of the phenotype of cancer cells that may contribute to tumour progression (7, 36). Elevated expression of c-*myc* has been shown to replace in part platelet-derived growth factor in induction of competence for DNA replication (7). Generally, enhanced expression of an oncogene could be a necessary prerequisite for acquisition of a growth advantage by cells having extra copies of the gene. This effect could also be the principal contribution of amplification to tumourigenesis.

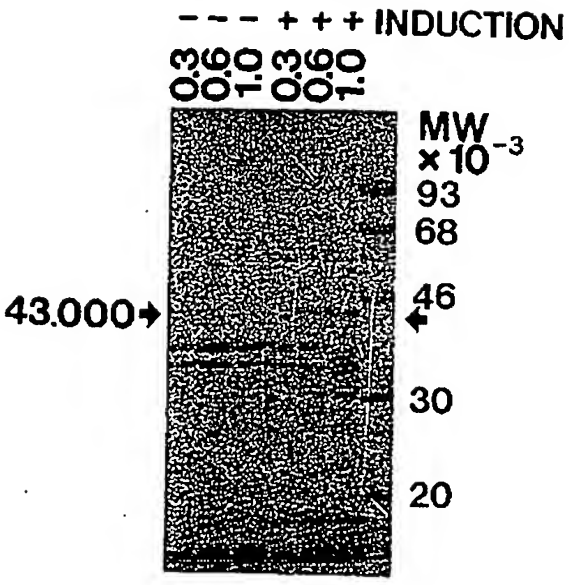


Fig. 10. *E. coli* 294 was transfected with the hybrid v-*src*, v-*myc* plasmid outlined in Fig. 9 and ampicillin-resistant bacterial colonies were checked for the production of a 43,000 m.w. bacterial v-*myc* protein after induction by growth to different optical densities in minimal essential medium (M9, induction +) or complete medium (LB, induction—) (3).

TUMOUR CELL AND STAGE SPECIFICITY OF ONCOGENE ACTIVATION AND AMPLIFICATION

Tumour cell specificity of oncogene amplification has been found in three malignancies. The c-*myc*, L-*myc* or N-*myc* oncogene is amplified in most cases of the variant form of small-cell lung cancer cells (53, 69), c-*erb*B is amplified in several glioblastomas (Josef Schlessinger, personal communication) and the putative N-*myc* oncogene is amplified in about half of grade III and IV neuroblastomas (14, 72, 73, 75). In addition to HSR:s, small-cell lung cancers and neuroblastomas frequently show a

Fig. 11. Indirect immunofluorescence for the v-*myc* protein and phase contrast microscopy of myelocytomatosis virus-transformed quail cells (3). A. Quail cells transformed with the MC-29 virus (Q8 cells). Anti-*myc* protein staining. B. Phase contrast microscopy of field in A. C. Q8 cells stained with anti-*myc* protein antiserum that has been blocked with the immunogen. D. Phase contrast microscopy of field in B. E. Q8 cell stained with preimmune rabbit serum. F. Phase contrast microscopy of field in E.

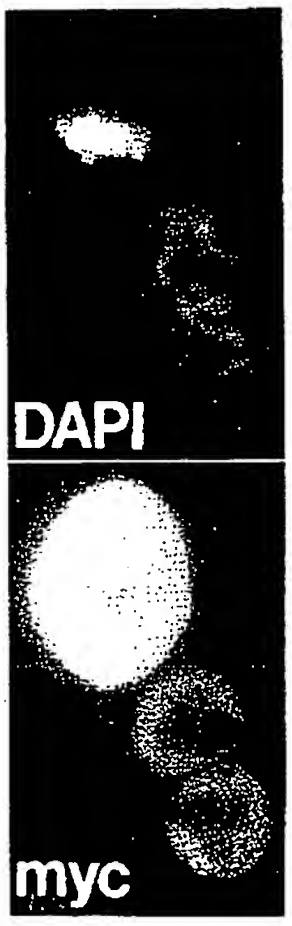


Fig. 12. Fluorescent staining for DNA and *myc* protein in myelocytomatosis virus-transformed quail cells. In interphase cells, the *myc* protein is confined to the nucleae. In the mitotic cell, *myc* fluorescence is distributed throughout the cell unlike fluorescence for chromatin, which is compacted to chromosomes in the metaphase plate. In fact, there is less *myc* fluorescence associated with chromatin than with the rest of the cell. DAPI, diaminophenylindole DNA stain. The anti-*myc* protein rabbit antiserum was used in a 1/200 dilution (ref. 89).

deletion of a portion of the short arm of chromosome 1 (13) and chromosome 3 (86, 87), respectively, in karyological examination. Two kinds of changes have also been described in different neuroblastoma oncogenes. The first is a mutation in the N-*ras* gene, an activated oncogene that was discovered because of its relation to other *ras* genes and transforming activity in transfection experiments (77). The second is amplification of a distant homologue of the c-*myc* gene called N-*myc* (72, 73, 75). Although the transforming potential of the N-*myc* gene has not yet been established, its consistent presence in a core segment of amplified neuroblastoma DNA (38, 57, 72, 73, 75) and its elevated expression in most retinoblastomas (48) suggests its oncogenic nature.

Taya et al. (80) have recently described a human lung giant cell carcinoma grown in nude mice, where both c-Ki-*ras* and c-*myc* oncogenes were amplified about 10-fold. Besides, sequencing studies indicated that at least some of the amplified c-Ki-*ras* copies were also mutationally activated in the 12th codon. These results fit to the multistage theory of cancer development and progression (see 58). Apparently co-operating lesions in cellular oncogenes accumulate during tumour growth and selection and increase the malignant potential of the tumour cells (44).

When does oncogene amplification come into play during tumourigenesis? Gene amplification may not be any initiating event in carcinogenesis. Amplification and enhanced expression of c-*myc* and N-*myc* may occur during the progression of human carcinoma of the lung and neuroblastoma cells to a more malignant phenotype (14, 53, 73). There may be, however, no mandatory sequence of oncogene amplifications for the genesis of any particular tumor. Amplification of an oncogene could play its part in malignant progression of already initiated cells whenever it happened to occur.

ACKNOWLEDGEMENTS

I thank my colleagues Manfred Schwab, Gerard Evan, J. Michael Bishop, Robert Winqvist, Kalle Saksela, Jorma Keski-Oja, C. C. Lin, Arthur Levinson, Wendy Colby and Donna George for collaboration, Ron Ellis for monoclonal antibodies against p21 proteins and Stephen Hann and Robert Eisenman for communicating their results before publication. The studies in the author's laboratory were supported by the Finnish Cancer Research Fund and by the Academy of Finland.

REFERENCES

1. *Abrams HD, Rohrschneider LR, Eisenman RN:* Nuclear location of the putative transforming protein of avian myelocytomatosis virus. Cell 29: 427–439, 1982
2. *Alitalo K, Bishop JM, Smith DH, Chen EY, Colby WW, Lenvinson AD:* Nucleotide sequence of the v-*myc* oncogene of avian retrovirus MC29. Proc Natl Acad Sci USA 80: 100–104, 1983
3. *Alitalo K, Ramsay G, Bishop JM, Ohlsson-Pfeifer S, Colby WW, Levinson AD:* Identification of nuclear proteins encoded by viral and cellular *myc* oncogenes. Nature 306: 274–277, 1983
4. *Alitalo K, Saksela K, Winqvist R, Schwab M, Bishop JM:* Amplification and aberrant expression of cellular oncogenes in human colon cancer cells. In: Genes and cancer. Ed. J. M. Bishop and J. Rowley. Alan Liss Co, New York, in press
5. *Alitalo K, Schwab M, Lin CC, Varmus HE, Bishop JM:* Homogeneously staining chromosomal regions con-

tain amplified copies of an abundantly expressed cellular oncogene (c-*myc*) in malignant neuroendocrine cells from a human colon carcinoma. Proc Natl Acad Sci USA 80: 1707—1711, 1983

6. *Alitalo K, Winqvist R, Lin CC, de la Chapelle A, Schwab M, Bishop JM:* Aberrant expression of an amplified c-*myb* oncogene in two cell lines from a colon carcinoma. Proc Natl Acad Sci USA 81: 4534—4538, 1984
7. *Armelin HA, Armelin MCS, Kelly K, Stewart T, Leder P, Cochran BH, Stiles CD:* Functional role for c-*myc* in mitogenic response to platelet-derived growth factor. Nature 310: 655—660, 1984
8. *Barker PE:* Double Minutes in human tumor cells. Cancer Genet Cytogenet 5: 81—94, 1982
9. *Barker PE, Drwinga HL, Mittelman WN, Maddox AM:* Double minutes replicate once during S phase of the cell cycle. Exp Cell Res 130: 353—360, 1980
10. *Barsoum J, Varshavsky A:* Mitogenic hormones and tumor promoters greatly increase the incidence of colony-forming cells bearing amplified dihydrofolate reductase genes. Proc Natl Acad Sci USA 80: 5330—5334, 1983
11. *Biedler JL, Meyers MB, Spengler BA:* Homogeneously staining regions and double minute chromosomes, prevalent cytogenetic abnormalities of human neuroblastoma cells. Adv Cell Neurobiol 4: 268—301, 1983
12. *Biedler JL, Spengler BA:* Metaphase chromosome anomaly: association with drug resistance and cell-specific products. Science 191: 185—187, 1976
13. *Brodeur GM, Green AA, Hayes FA, Williams KJ, Williams DL, Tsiatis AA:* Cytogenetic features of human neuroblastomas and cell lines. Cancer Res 41: 4678—4686, 1981
14. *Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM:* Amplification of N-*myc* in Untreated Human Neuroblastomas Correlates with Advanced Disease Stage. Science 224: 1121—1124, 1984
15. *Brown PC, Tlsty TD, Schimke RT:* Enhancement of methotrexate resistance and DHFR gene amplification by treatment of 3T6 cells with hydroxyurea. Mol Cell Biol 3: 1097—1107, 1983
16. *Bulloch P, Botchan M:* Molecular events in the excicion of SV40 DNA from the chromosomes of cultured mammalian cells. In: Gene amplification. Ed. R. T. Schimke. Cold Spring Harbor Lab 215—230, 1982
17. *Calabretta B, Robberson DL, Barrera-Saldana HA, Lambrou TP, Saunders GF:* Genome instability in a region of human DNA enriched in Alu repeat sequences. Nature 296: 219—225, 1982
18. *Chattopadhyay SK, Chang EH, Lander MR, Ellis RW, Scolnick EM, Lowy DR:* Amplification and rearrangement of *onc* genes in mammalian species. Nature 296: 361—363, 1982
19. *Cifone MA, Fidler IJ:* Increasing metastatic potential is associated with increasing genetic instability of clones isolated from murine neoplasms. Proc Natl Acad Sci USA 78: 6949—6952, 1981
20. *Collins S, Groudine M:* Amplification of endogenous *myc*-related DNA sequences in a human myeloid leukemia cell line. Nature 298: 679—681, 1982
21. *Collins SJ, Groudine MT:* Rearrangement and amplification of c-*abl* sequences in the human chronic myelogenous leukemia cell line K-562. Proc Natl Acad Sci USA 80: 4813—4817, 1983
22. *Collins SJ, Kubonishi I, Miyoshi I, Groudine MT:* Altered transcription of the c-*abl* oncogene in K-562 and other chronic myelogenous leukemia cells. Science 225: 72—74, 1984
23. *Cowell JK:* Double minutes and homogeneously staining regions: Gene amplification in mammalian cells. Ann Rev Genet 16: 21—52, 1982
24. *Cox D, Yuncken C, Spriggs AI:* Numerous minute chromatin bodies in malignant tumours of childhood. Lancet I: 55—58, 1965
25. *Dalla-Favera RD, Wong-Staal F, Gallo RG: Onc* gene amplification in promyelocytic leukemia cell line HL-60 and primary leukaemic cells of the same patient. Nature 299: 61—63, 1982
26. *Donner P, Greiser-Wilke I, Moelling K:* Nuclear Localization and DNA binding of the transforming gene product of avian myelocytomatosis virus. Nature 296: 262—266, 1982
27. *Downward J, Yarden Y, Mayes E, Scrace G, Totty N, Stockwell P, Ullrich A, Schlessinger J, Waterfield MD:* Close similarity of epidermal growth factor receptor and v-*erb*B oncogene protein sequences. Nature 307: 521—527, 1984
28. *Echols H:* SOS-functions, cancer and inducible evolution. Cell 25: 1—2, 1981
29. *Eisenman RN, Tachibana CY, Abrams HD, Hann SR:* v-*myc* and c-*myc*-Encoded proteins are associated with the nuclear matrix. Submitted
30. *Gilbert F, Balaban G, Brangman D, Herrmann N, Lister A:* Homogeneously staining regions and tumorigenicity. Int J Cancer 31: 765—768, 1983
31. *Gill GN, Lazar CS:* Increased phosphotyrosine content and inhibition of proliferation in EGF-treated A431 cells. Nature 293: 305—307, 1981
32. *Hann SR, Abrams HD, Rohrschneider LR, Eisenman RN:* Proteins encoded by v-*myc* and c-*myc* oncogenes: Identification and localization in acute leukemia virus transformants and bursal lymphoma cell lines. Cell 34: 789—798, 1983
33. *Hann SR, Eisenman RN:* Two proteins encoded by the human c-*myc* oncogene: Differential expression in neoplastic cells. Mol Cell Biol 4: 2486—2497, 1984
34. *Harper ME, Franchini G, Love J, Simon MI, Gallo RC, Wong-Staal F:* Chromosomal sublocalization of human c-*myb* and c-*fes* cellular onc genes. Nature 304: 169—171, 1983
35. *Heisterkamp N, Stephenson JR, Groffen J, Hansen PF, de Klein A, Bartman CR, Grosveld G:* Localization of the c-*abl* oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. Nature 306: 239—242, 1983
36. *Heldin C-H, Westermark B:* Growth factors: Mechanism of action and relation to oncogenes. Cell 37: 9—20, 1984
37. *Johnston RN, Beverley SM, Schimke RT:* Rapid spontaneous dihydrofolate reductase gene amplification shown by fluorescence-activated cell sorting. Proc Natl Acad Sci USA 80: 3711—3715, 1983
38. *Kanda N, Schreck R, Alt F, Bruns G, Baltimore D, Latt S:* Isolation of amplified DNA sequences from IMR-32 human neuroblastoma cells: Facilitation by fluoroscence-activated flow sorting of metaphase chromosomes. Proc Natl Acad Sci USA 80: 4069—4073, 1983
39. *Kelly K, Cochran BH, Stiles CD, Leder P:* Cell-specific regulation of the c-*myc* gene by lymphocyte mitogens and platelet-derived growth factor. Cell 35: 603—610, 1983
40. *Klein G:* Specific chromosome translocations and the genesis of B-cell derived tumors in mice and men. Cell 32: 311—315, 1983
41. *Konopka JB, Watanabe SM, Witte ON:* An alteration of the human c-*abl* protein in K562 leukemia cells unmasks associated tyrosine kinase activity. Cell 37: 1035—1042, 1984
42. *Kovacs G:* Homogeneously staining regions on marker

chromosomes in malignancy. Int J Cancer 23: 299—301, 1979
43. *Kozbor D, Croce CM:* Amplification of the c-*myc* oncogene in one of five human breast carcinoma cell lines. Cancer Res 44: 438—441, 1984
44. *Land H, Parada LF, Weinberg RA:* Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Science 222: 771—778, 1983
45. *Latt SA:* Sister chromatid exchange formation. Ann Rev Genet 15: 11—15, 1981
46. *Laughlin TJ, Taylor JH:* Initiation of DNA replication in chromosomes of Chinese hamster ovary cells. Chromosoma 75: 19—35, 1979
47. *Lavi S:* Carcinogen-mediated amplification of viral DNA sequences in simian virus 40-transformed Chinese hamster embryo cells. Proc Natl Acad Sci USA 78: 6144—6148, 1981
48. *Lee W-H, Murphree AL, Benedict WF:* Expression and amplification of the N-*myc* gene in primary retinoblastoma. Nature 309: 458—460, 1984
49. *Levan A, Manolov G, Clifford P:* Chromosomes of a human neuroblastoma. A new case with accessory minute chromosomes. J Natl Cancer Inst 41: 1377—1387, 1968
50. *Levan A, Levan G, Mitelman F:* Chromosomes and cancer. Hereditas 86: 15—29, 1977
51. *Lewis JA, Biedler JL, Melera PW:* Gene amplification accompanies low level increases in the activity of dihydrofolate reductase in antifolate-resistant Chinese hamster lung cells containing abnormally banding chromosomes. J Cell Biol 94: 418—424, 1982
52. *Lin CC, Alitalo K, Schwab M, George D, Varmus HE, Bishop JM:* Evolution of karyotypic abnormalities and c-*myc* oncogene amplification in a human colonic carcinoma. Submitted for publication
53. *Little CD, Nau MM, Carney DN, Gazdar AF, Minna JD:* Amplification an expression of the c-*myc* oncogene in human lung cancer cell lines. Nature 306: 194—196, 1983
54. *Lozzio CB, Lozzio BB:* Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood 45: 321—334, 1975
55. *Mariani BD, Schimke RT:* Gene amplification in a single cell cycle in Chinese hamster ovary cells. J Biol Chem 259: 1901—1910, 1984
56. *Mark J:* Double minutes-a chromosomal aberration in Rous sarcomas in mice. Hereditas 57: 1—22, 1967
57. *Montgomery KT, Biedler JL, Spengler BA, Melera PW:* Specific DNA sequence amplification in human neuroblastoma cells. Proc Natl Acad Sci USA 80: 5724—5728, 1983
58. *Nowell RC:* The clonal evolution of tumor cell populations. Science 194: 23—28, 1976
59. *Nowell R, Finan J, Favera RD, Gallo RC, Ar-Rushdi A, Romanczuk G, Selden JR, Emanuel BS, Rovera G, Croce CM:* Association of amplified oncogene c-*myc* with an abnormally banded chromosome 8 in a human leukemia cell line. Nature 306: 494—497, 1983
60. *Pall ML:* Gene amplification model of carcinogenesis. Proc Natl Acad Sci USA 78: 2465—2468, 1981
61. *Pelicci P-G, Lanfrancone L, Brathwaite MD, Wolman SR, Dalla-Favera R:* Amplification of the c-*myb* oncogene in a case of human acute myelogenous leukemia. Science 224: 1117—1121, 1984
62. *Pellegrini S, Dailey L, Basilico C:* Amplification and excision of integrated polyoma DNA sequences require a functional origin of replication. Cell. 36: 943—949, 1984
63. *Quinn LA, Moore GE, Morgan RT, Woods LK:* Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions. Cancer Res 39: 4914, 1979
64. *Rechavi G, Givol D, Canaani E:* Activation of a cellular oncogene by DNA rearrangement: possible involvement of an IS-like element Nature 300: 607—611, 1982
65. *Roberts JM, Buck LB, Axel R:* A structure for amplified DNA. Cell 33: 53—63, 1983
66. *Rowley JD:* Human oncogene locations and chromosome aberrations. Nature 301: 290, 1983
67. *Rowley JD, Testa JR:* Chromosome abnormalities in malignant hematologic diseases. Adv Cancer Res 36: 103—147, 1982
68. *Rowley JD, Ultmann JE eds:* Chromosomes and Cancer: from molecules to man. Academic Press, New York 1983
69. *Saksela K, Bergh J, Lehto V-P, Nilsson K, Alitalo K:* Amplification of the c-*myc* oncogene is characteristic of a subpopulation of human small cell lung cancer. Cancer Res, in press
70. *Schimke RT:* Gene amplification. Cold Spring Harbor Labor 1982
71. *Schimke RT, Brown PC, Kaufman RJ, McGrogan M, Slate DL:* Chromosomal and extrachromosomal localization of amplified dihydrofolate reductase genes in cultured mammalian cells. Cold Spring Harbor Symp Quant Biol 55: 785—797, 1981
72. *Schwab M, Alitalo K, Klempnauer K-H, Varmus HE, Bishop JM, Gelbert F, Brodeur G, Goldstein M, Trent J:* Amplified DNA with limited homology to *myc* cellular oncogne is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature 305: 245—248, 1983
73. *Schwab M, Alitalo K, Varmus HE, Bishop JM:* Amplification of cellular oncogenes in tumor cells. In: The cancer cell. Cold Spring Harbor Press 2: 215—220, 1984
74. *Schwab M, Alitalo K, Varmus HE, Bishop JM, George D:* A cellular oncogene [c-Ki-*ras*] is amplified, overexpressed, and located within karyotypic abnormalities in mouse adreocortical tumour cells. Nature 303: 497—501, 1983
75. *Schwab M, Varmus HE, Bishop JM, Grezeschik K-H, Naylor SL, Sakaguchi AY, Brodeur G, Trent J:* Chromosome localization in normal human cells and neuroblastomas of a gene related to c-*myc*. Nature 308: 288—291, 1984
76. *Selden JR, Emanuel BS, Wang E, Cannizzaro L, Palumbo A, Erikson J, Nowell PC, Rovera G, Croce CM:* Amplified C_λ and c-*abl* genes are on the same marker chromosome in K562 leukemia cells. Proc Natl Acad Sci USA 80: 7289—7292, 1983
77. *Shimizu K, Goldfarb M, Perucho M, Wigler M:* Isolation and preliminary characterization of the transforming gene of human neuroblastoma cell line. Proc Natl Acad Sci USA 80: 383—387, 1983
78. *Shmookler Reis RJ, Lumpkin CK, McGill JR, Riabowol TK, Goldstein S:* Extrachromosomal circular copies of an 'inter-Alu' unstable sequence in human DNA are amplified during in vitro and in vivo ageing. Nature 201: 394—398, 1983
79. *Stark GR, Wahl GM:* Gene amplification. Ann Rev Biochem 53: 447—491, 1984
80. *Taya Y, Hesogai K, Hirohashi S, Shimosato Y, Tsuchiya R, Tsuchida N, Fushimi M, Sekiya T, Nishimura S:* A novel combination of K-*ras* and *myc* amplification accompanied by point mutational activation of K-*ras* in human lung cancer. EMBO J 3: 2943—2946, 1984
81. *Tlsty TD, Brown PC, Schimke TR:* UV radiation facilitates methotrexate resistance and amplification of the dihydrofolate reductase gene in cultured 3T6 mouse cells. Molec Cell Biol 4: 1050—1056, 1984

82. *Ullrich A, Coussens L, Hayklick JS, Dull TJ, Gray A, Tam AW, Lee J, Yarden Y, Libermann TA, Sschlessinger J, Downward J, Mayes ELV, Whittle N, Waterfield MD, Seeburg PH:* Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature 309: 418—425, 1984
83. *Wahl GM, de Saint Vincent BR, DeRose ML:* Effect of chromosomal position on amplification of transfected genes in animal cells. Nature 307: 516—520, 1984
84. *Varshavsky A:* On the possibility of metabolic control of replicon 'misfiring': Relationship to emergence of malignant phenotypes in mammalian cell lineages. Proc Natl Acad Sci USA 78: 3673—3677, 1981
85. *Varshavsky A:* Phorbol ester dramatically increases incidence of methotrexate-resistant mouse cells: Possible mechanisms and relevance to tumor promotion. Cell 25: 561—572, 1981
86. *Whang-Peng J, Bunn PA, Kao-Shan CS, Lee EC, Carney DN, Gazdar A, Minna JD:* A Nonrandom chromosomal abnormality, del 3p (14—23), in human small cell lung cancer (SCLC). Cancer Genet Cytogenet 6: 119—134, 1982
87. *Whang-Peng J, Kao-Shan CS, Lee EC, Bunn PA, Carney DN, Gazdar AF, Portlock C, Minna JD:* Deletion 3p (14—23), Double minute chromosomes, and homogeneously staining regions in human small-cell lung cancer. sCSH Laboratory 1982
88. *Winqvist R, Knuutila S, Leprince D, Stehelin D, Alitalo K:* Mapping of amplified c-*myb* oncogene, sister chromatid exchanges and karyotypic analysis of the COLO 205 colon carcinoma cell line. Cancer Genet Cytogenet. in press
89. *Winqvist R, Sahsela K, Alitalo K: myc* proteins are not associated with chromatin in mitotic cells. EMBO J 3: 2947—2950
90. *Woodcock DM, Cooper IA:* Evidence for double replication of chromosomal DNA segments as a general consequence of DNA replication inhibition. Cancer Res 41: 2483—2490, 1981
91. *Zabel BU, Naylor SL, Grezeschik K-H, Sakaguchi AY:* Regional assignment of human protooncogene c-*myb* to 6q21—qter. Somatic Cell Molec Genet 10: 105—108, 1984

Received for publication: September 19, 1984
Address: K. Alitalo
Department of Virology
University of Helsinki
SF-00290 Helsinki
Finland

Elevated Epidermal Growth Factor Receptor Gene Copy Number and Expression in a Squamous Carcinoma Cell Line

Glenn T. Merlino, Young-hua Xu, Nancy Richert, Adrian J. L. Clark, Shunsuke Ishii, Susan Banks-Schlegel, and Ira Pastan

Laboratory of Molecular Biology, Division of Cancer Biology and Diagnosis, and Laboratory of Human Carcinogenesis, Division of Cancer Etiology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Abstract

The human epidermal growth factor (EGF) receptor is known to be homologous to the v-*erb* B oncogene protein of the avian erythroblastosis virus. Overexpression of the EGF receptor gene in A431 epidermoid carcinoma cells is due to gene amplification. In this study, a variety of squamous cell carcinomas were examined and one, SCC-15, contained high levels of the EGF receptor as determined by immunoprecipitation via an EGF receptor-specific polyclonal antibody. Using a cloned EGF receptor complementary DNA as a probe, the level of EGF receptor RNA was found to be elevated four-fold in SCC-15 relative to normal cultured keratinocytes. When the same probe was used to identify EGF receptor gene fragments on a genomic DNA blot, the SCC-15 cell line was shown to possess an EGF receptor gene copy number amplified four to five times. Gene amplification results in the enhancement in the level of the EGF receptor in several carcinomas and could be responsible for the appearance of the transformed phenotype in these cells.

Introduction

The epidermal growth factor (EGF)[1] stimulates growth and elicits a wide variety of rapid and delayed responses by binding to high-affinity cell-surface receptors which are 170-kD glycoproteins (1). Recently, EGF receptor peptides have been sequenced and found to be homologous to the avian erythroblastosis virus *erb* B oncogene product (2), suggesting that the EGF receptor gene is the human c-*erb* B oncogene. A431 epidermoid carcinoma cells possess a very large number of EGF receptors (3), and the EGF receptor gene is amplified ~30-fold (4–6). This amplification is responsible for the overexpression of the EGF receptor protein in these cells (4–6).

A cell culture system has been developed permitting serial cultivation of keratinocytes, whose growth is modulated by EGF (7). Such methods have been used to establish cell lines from squamous cell carcinomas of the oral epithelium (8). Because of the role of EGF in keratinocyte development, we quantified EGF receptor protein and RNA in several squamous cell carcinomas. One cell line, SCC-15, was found to contain high amounts of receptor protein and RNA, and a four- to fivefold amplification of the gene.

Methods

The squamous cell carcinomas established by Rheinwald and Beckett (8) were obtained from, and maintained according to the American Type Culture Collection (Rockville, MD). 1623 was originally designated as SCC-15; 1628 as SCC-25; and 1629 as SCC-9 (8). Normal human esophageal epithelial cells were grown as reported (9). Maintenance of other cell lines was as described elsewhere (10). Proteins were labeled with [^{35}S]methionine and immunoprecipitated as previously described (10). PolyA+ RNA was isolated by guanidine isothiocyanate solubilization and CsCl centrifugation (11), and oligo(dT)-affinity chromatography. RNA (Northern) blotting was performed as described (11, 12). High molecular weight DNA was isolated (4) and analyzed by DNA (Southern) blotting (4, 10, 13). The EGF receptor complementary DNA (cDNA) clone pE7 was constructed and isolated from an A431 cDNA library (11). DNA fragments were ^{32}P-labeled by nick translation.

Results

A large number of cell lines were initially screened for EGF receptor levels by determining their ability to be killed by an EGF-pseudomonas exotoxin conjugate, a technique described previously (10). Several squamous cell carcinomas were found to be relatively sensitive to the EGF-toxin conjugate, including SCC-25, SCC-9, and particularly SCC-15, all derived from the human tongue (8). These three cell lines were labeled with [^{35}S]methionine, and their extracts immunoprecipitated with a goat polyclonal antibody to the EGF receptor, affinity-purified as described (10). When compared with A431 cells, which make very large amounts of the EGF receptor, SCC-25 and SCC-9 make moderate amounts and SCC-15 high amounts of the receptor (Fig. 1, lane *a* vs. *e*, *g*, and *c*). Quantitation of the immunoprecipitation data revealed that SCC-15, SCC-25, and SCC-9 make 41, 15, and 4% of the amount of EGF receptor made by A431 cells, respectively.

Because SCC-15 cells had high levels of receptor, polyA+ RNA was isolated from these cells, electrophoretically fractionated on agarose, and analyzed by RNA (Northern) blotting. A cloned A431 cDNA (pE7) encoding the EGF receptor (11) was ^{32}P-labeled and used as a hybridization probe to visualize EGF receptor RNAs. Fig. 2 *A* shows that SCC-15 contains both the 10- and 5.6-kilobase species of EGF receptor RNA. The levels are approximately four- to fivefold higher than those found in either KB or A498 kidney carcinoma cells; these cell lines were previously found to possess readily de-

Received for publication 21 November 1984.

1. *Abbreviations used in this paper:* cDNA, complementary DNA; EGF, epidermal growth factor.

The Journal of Clinical Investigation, Inc.
Volume 75, March 1985, 1077–1079

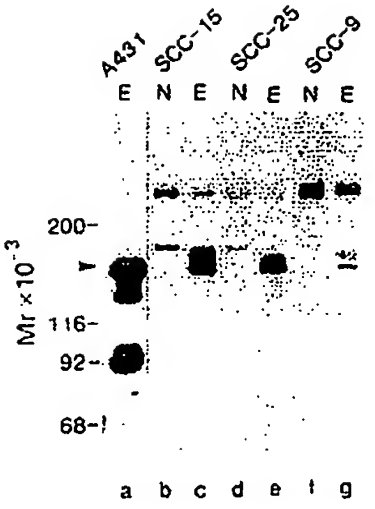


Figure 1. Autoradiograph of electrophoresed ^{35}S-labeled EGF receptor protein (arrow) immunoprecipitated by either E, affinity-purified goat polyclonal EGF receptor antibody, or N, normal serum. Molecular weight markers are at left.

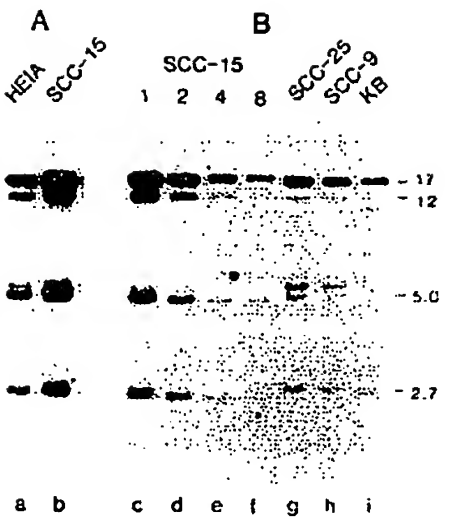


Figure 3. DNA blot analysis of *Hin*dIII-digested genomic DNAs using the pE7 probe (see Fig. 2). (*A*) and (*B*) are autoradiographs from two separate gels. Sizes are in kilobase pairs (right). 10 μg of DNA was loaded except in (*B*), lanes *d–f*, which represent serial dilutions of the 1623 DNA shown in lane *c*.

tectable levels of both receptor RNAs (10). Fig. 2 *B* shows that cultured human epithelial cells contain EGF receptor-specific RNA (HEIA, lane *e*) whose levels are higher than an early passage human fibroblast D551 (lane *f*), equivalent to A498 (lane *g*), but much lower than SCC-15 (lane *b*).

To determine if an elevated gene copy number was associated with enhanced expression of the EGF receptor gene in SCC-15 cells, genomic DNA was isolated from normal cultured epithelial cells (HEIA) and SCC-15 cells, digested with *Hin*dIII, electrophoretically fractionated, and subjected to DNA blotting analysis. An EGF receptor cDNA (pE7) was used as a hybridization probe to identify receptor DNA fragments. Fig. 3 *A* reveals that the SCC-15 genome contains four- to fivefold amplified EGF receptor gene sequences relative to normal epithelial cells (lane *a* vs. *b*). Analysis of β-actin gene fragments on the same filter by hybridization to a chick actin cDNA probe indicated that equal amounts of DNA were loaded per well (data not shown). Digested SCC-15 DNA had to be diluted about fourfold (Fig. 3 *B*, lane *e*) to approximate the signal intensity of receptor DNA fragments from SCC-25, SCC-9, and KB cells (lanes *g–i*). The KB cell EGF receptor gene is known not to be amplified (10).

Discussion

It may be significant that A431 carcinoma cells are not unique in their possession of amplified EGF receptor genes. We report here that the EGF receptor gene in squamous cell carcinoma SCC-15 is amplified four- to fivefold relative to normal epithelial cells. The amplification of the EGF receptor gene may cause the initiation or maintenance of the malignant state in some human cells.

Previously, we reported that a variety of transformed cell lines synthesize relatively high amounts of both EGF receptor protein and messenger RNA (10). It is conceivable that a moderate or even a small increase in the level of the EGF receptor leads to a change in the cellular phenotype, as has been demonstrated for the *src* gene product (14). If this hypothesis is correct, then even a minor amplification of the EGF receptor gene copy number could contribute to the onset of tumorigenesis. Hendler and Ozanne (15) have examined lung squamous cell carcinomas and found that they contain a 2.5–5-fold increase in EGF receptor levels.

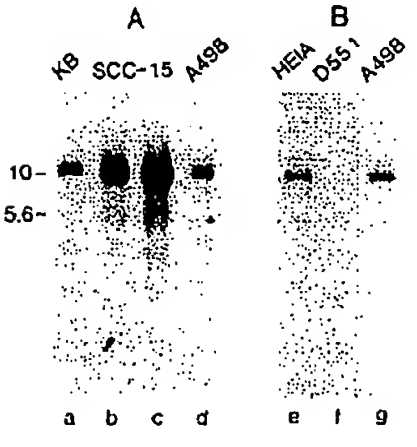


Figure 2. RNA blot analysis of polyA+ RNAs using the ^{32}P-labeled EGF receptor cDNA probe pE7. (*A*) and (*B*) are autoradiographs from two separate gels. Sizes are in kilobases (left). 5 (lanes *a*, *b*, *d*, and *g*) or 10 (lanes *c*, *e*, and *f*) μg of RNA were loaded.

References

1. Savage, C. R., and S. J. Cohen. 1972. Epidermal growth factor and a new derivation: rapid isolation procedure and biological and chemical characterization. *J. Biol. Chem.* 247:7609–7611.

2. Downward, J., Y. Yarden, E. Mayes, G. Scarce, N. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, and M. D. Waterfield. 1984. Close similarity of the epidermal growth factor receptor and v-erb B oncogene protein sequence. *Nature (Lond.)*. 307:521–527.

3. Fabricant, R. N., J. E. DeLarco, and G. J. Todaro. 1977. Nerve growth factor receptor on human melanoma cells in culture. *Proc. Natl. Acad. Sci. USA*. 74:565–569.

4. Merlino, G. T., Y.-H. Xu, S. Ishii, A. J. L. Clark, K. Semba, K. Toyoshima, T. Yamamoto, and I. Pastan. 1984. Amplification and enhanced expression of the epidermal growth factor receptor gene in A431 human carcinoma cells. *Science (Wash. DC)*. 224:417–419.

5. Ullrich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermal carcinoma cells. *Nature (Lond.)*. 309:418–425.

6. Lin, C. R., W. S. Chen, W. Kruiger, L. S. Stolarsky, W. Weber, R. M. Evans, I. M. Verma, G. N. Gill, and M. G. Rosenfeld. 1984.

Expression cloning of human EGF receptor complementary DNA: gene amplification and three related messenger RNA products in A431 cells. *Science (Wash. DC)*. 224:843–848.

7. Rheinwald, J. G., and H. Green. 1975. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell*. 6:331–344.

8. Rheinwald, J. G., and M. A. Beckett. 1981. Tumorigenic keratinocyte lines requiring anchorage and fibroblast support cultured from human squamous cell carcinomas. *Cancer Res.* 41:1657–1663.

9. Banks-Schlegel, S. P., and C. C. Harris. 1983. Tissue-specific expression of keratin proteins in human esophageal and epidermal epithelium and their cultured keratinocytes. *Exp. Cell Res.* 146:271–280.

10. Xu, Y.-H., N. Richert, S. Ito, G. T. Merlino, and I. Pastan. 1984. Characterization of epidermal growth factor receptor gene expression in malignant and normal human cell lines. *Proc. Natl. Acad. Sci. USA*. 81:7308–7312.

11. Xu, Y.-H., J. Ishii, A. J. L. Clark, M. Sullivan, R. K. Wilson, D. P. Ma, B. A. Roe, G. T. Merlino, and I. Pastan. 1984. Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs overproduced in A431 carcinoma cells. *Nature (Lond.)*. 309: 806–810.

12. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA*. 77:5201–5205.

13. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.

14. Jakobovits, E. B., J. E. Majors, and H. E. Varmus. 1984. Hormonal regulation of the Rous sarcoma src gene via a heterologous promoter defines a threshold dose for cellular transformation. *Cell*. 38: 757–765.

15. Hendler, F. J., and B. W. Ozanne. 1984. Human squamous cell lung cancers express increased epidermal growth factor receptors. *J. Clin. Invest.* 74:647–651.

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman‖, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. ***Molecular & Cellular Proteomics 1:37–45, 2002.***

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1*, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)[1] have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q−, 11p−, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

From the ‡Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and **Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ¶AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ‖UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

[1] The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

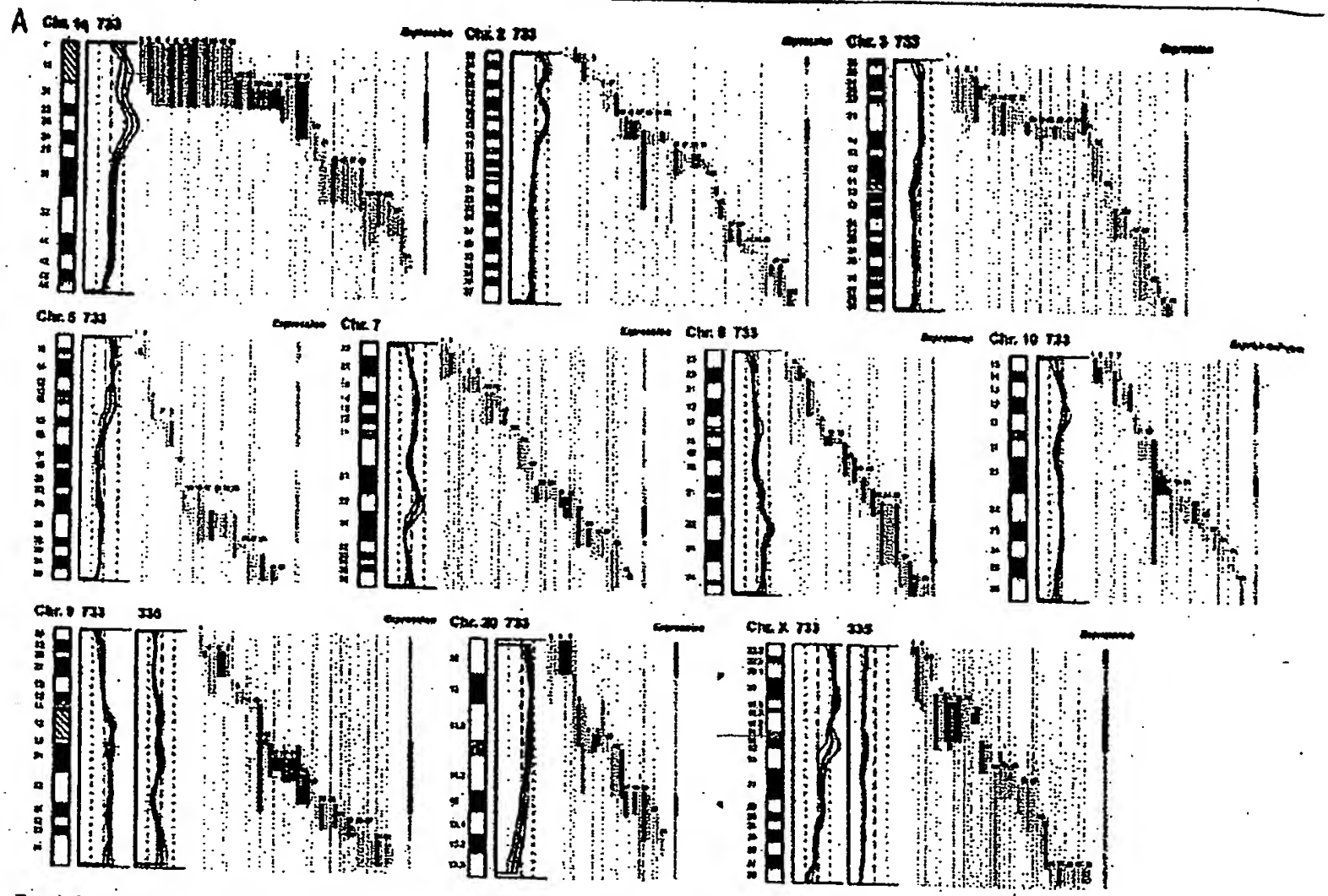


FIG. 1. **DNA copy number and mRNA expression level.** Shown from *left* to *right* are chromosome (*Chr.*), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. *A*, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. *B*, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (*left*). The *bold curve* in the ratio profile represents a mean of four chromosomes and is surrounded by *thin curves* indicating one standard deviation. The *central vertical line* (*broken*) indicates a ratio value of 1 (no change), and the *vertical lines* next to it (*dotted*) indicate a ratio of 0.5 (*left*) and 2.0 (*right*). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the *right* of the invasive tumor profile. The *colored bars* represents one gene each, identified by the running *numbers above* the *bars* (the name of the gene can be seen at www.MDL.DK/sdata.html). The *bars* indicate the purported location of the gene, and the *colors* indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (*black*), >2-fold decrease (*blue*), no significant change (*orange*). The *bar* to the *far right*, entitled *Expression* shows the resulting change in expression along the chromosome; the *colors* indicate that at least half of the genes were up-regulated (*black*), at least half of the genes down-regulated (*blue*), or more than half of the genes are unchanged (*orange*). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at −80 °C. Total RNA was isolated using the RNAzol B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)$^+$ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript® choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the MEGAscrip® *in vitro* transcription kit (Ambion). Biotin-labeled CTP and UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94 °C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6× SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95 °C for 5 min, subsequently cooled to 40 °C, and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40 °C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6× SSPE-T at 25 °C followed by 4 washes in 0.5× SSPE-T at 50 °C. The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 μg/ml (Molecular Probes) in 6× SSPE-T

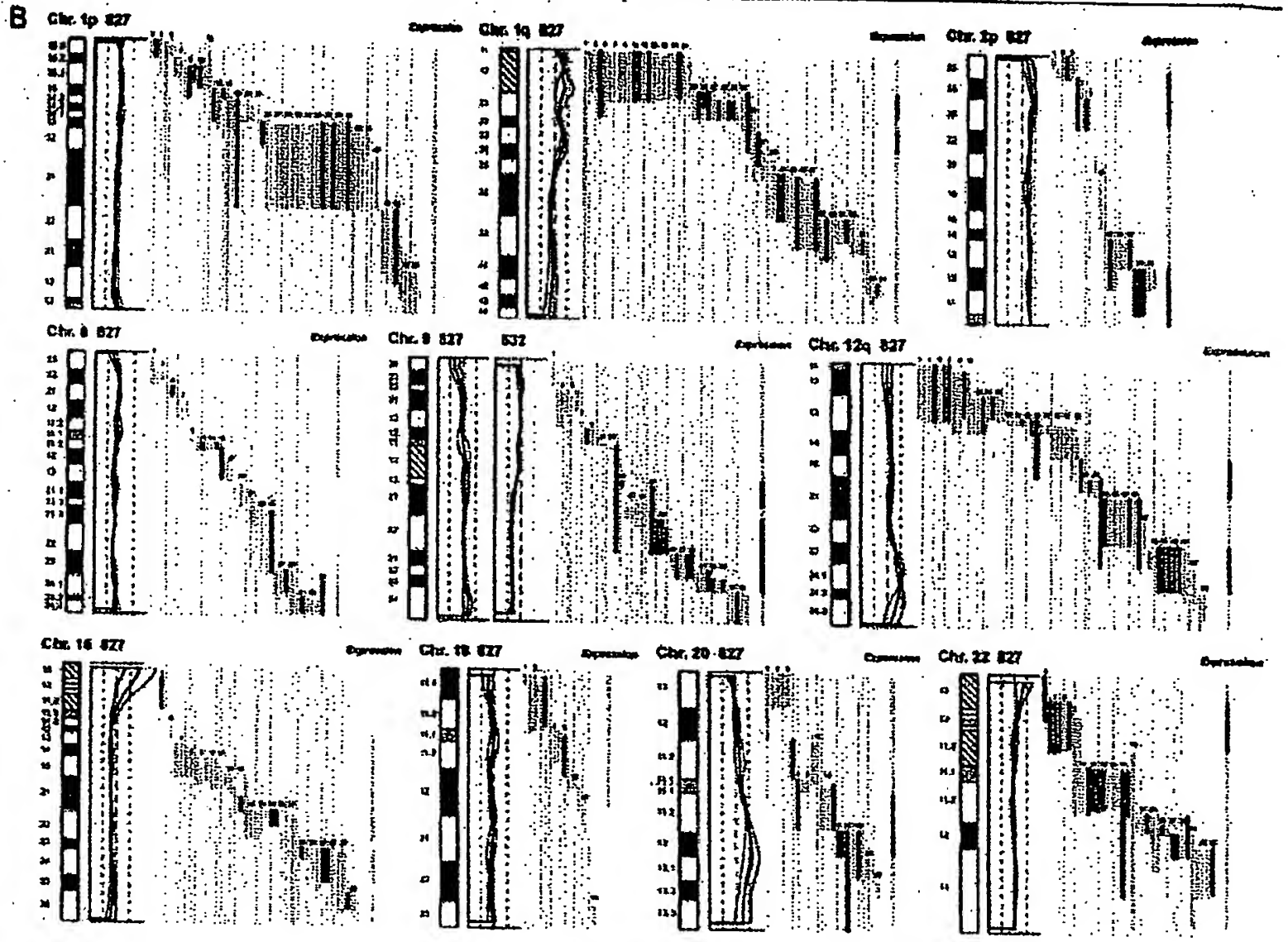


FIG. 1—*continued*

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μl for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at −20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/celis.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-labeled reference DNA (200 ng), and human Cot-1 DNA (20 μg) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μg/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

TABLE I

Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%

Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22-q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22-24+, 2q14.1-qter–, 3q12-q13.3–, 6q12-q22–, 9q34+, 11q12-q13+, 17+, and 20q11.2-q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1*A*) and 20q12 in TCC 827 (Fig. 1*B*).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21-q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, *top*). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, *top*). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, *bottom*). Furthermore, as a control we looked at areas with no alter-

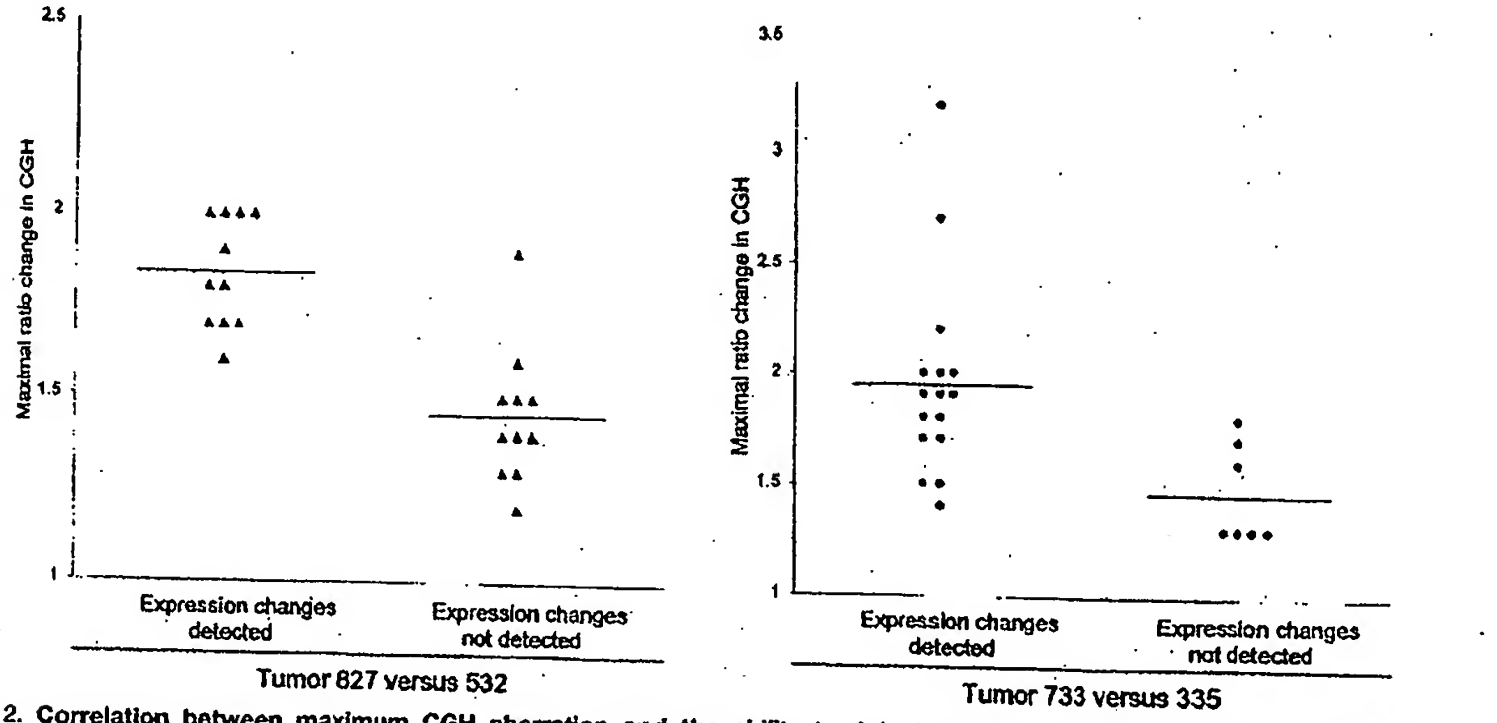


FIG. 2. **Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring.** The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (♦) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the *right* in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1*A*). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

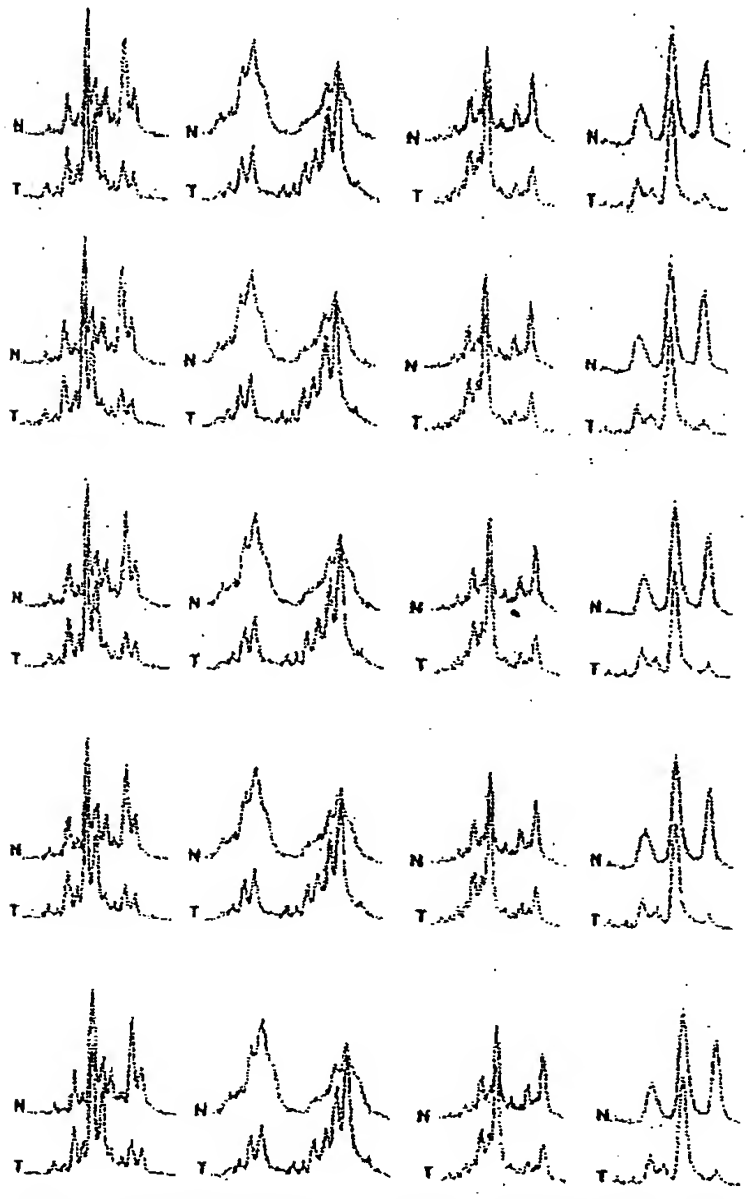


FIG. 3. **Microsatellite analysis of loss of heterozygosity.** Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (*a*) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (*b*) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (*c*) at chromosome 2p23 by D2S2251 close to general β-spectrin (gene number 11 on Fig. 1) and of (*d*) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The *upper curves* show the electropherogram obtained from normal DNA from leukocytes (*N*), and the *lower curves* show the electropherogram from tumor DNA (*T*). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH

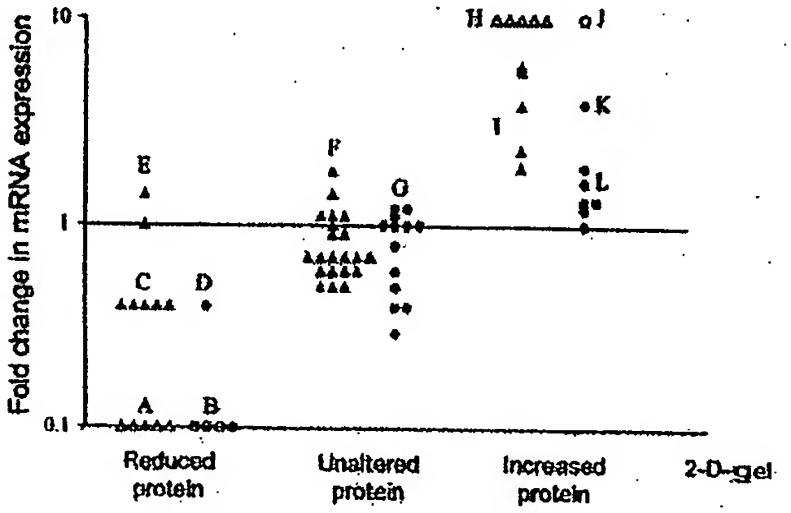


FIG. 4. **Correlation between protein levels as judged by 2D-PAGE and transcript ratio.** For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (*horizontal axis*). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (*vertical axis*). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; Δ, mRNAs that were scored as absent in the invasive tumors (along *horizontal axis*) or as absent in non-invasive reference (*top* of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲Δ) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group *A* (from *left*), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytokeratin 15, and cytokeratin 13; *B* (from *left*), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytokeratin 13, and calcyclin; *C* (from *left*), α-enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3-ε, and pre-mRNA splicing factor; *D*, mesothelial keratin K7 (type II); *E* (from *top*), glutathione *S*-transferase-π and mesothelial keratin K7 (type II); *F* (from *top* and *left*), adenylyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ-actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β-1 subunit; *G*, (from *top* and *left*), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytokeratin 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione *S*-transferase-π, and keratin 8; *H* (from *left*), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15 hydroxyprostaglandin dehydrogenase; *I* (from *top*), prolyl 4-hydroxylase β-subunit, cytokeratin 20, cytokeratin 17, prohibitin, and fructose 1,6-biphosphatase; *J* annexin II; *K*, annexin IV; *L* (from *top* and *left*), 90-kDa heat shock protein, prolyl 4-hydroxylase β-subunit, α-enolase, GRP 78, cyclophilin, and cofilin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

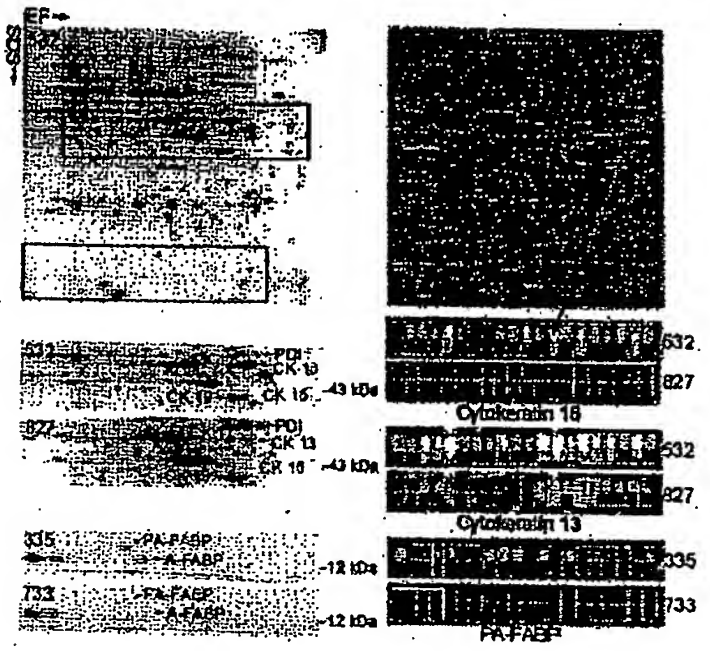


FIG. 5. **Comparison of protein and transcript levels in invasive and non-invasive TCCs. The *upper part* of the figure shows a 2D gel (*left*) and the oligonucleotide array (*right*) of TCC 532. The *red rectangles* on the *upper gel* highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (*red annotation*). The tile on the array containing probes for cytokeratin 15 is enlarged *below* the array (*red arrow*) from TCC 532 and is compared with TCC 827. The *upper row* of *squares* in each tile corresponds to perfect match probes; the *lower row* corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as *black*, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the *bottom* of the figure (*left*) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the *right* we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). *IEF*, isoelectric focusing.**

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a know chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II

Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration[a]	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres[a]	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up[b]	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

[a] Abs, absent; Pres, present.

[b] In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p−, 9q−, 1q+, Y− (2, 6), and in pT1 tumors, 2q−,11p−, 11q−, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p− and 9q22-q33− and 9q− and Y−, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p−, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicate that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Vikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: ornoft@kba.sks.au.dk.

REFERENCES

1. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* **17**, 643–649
2. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46
3. de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Poulillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* **83**, 157–161
4. Brungler, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors: relationship with chromosome 11q13 amplifications. *Oncogene* **12**, 1747–1753
5. Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* **50**, 1459–1463
6. Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) c-*myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am J. Pathol.* **146**, 1131–1139
7. Richter, J., Jiang, F., Gorog, J. P., Sartorious, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* **57**, 2860–2864
8. Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* **153**, 1615–1621
9. Bruch, J., Wohr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* **23**, 167–174
10. Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* **15**, 3555–3560
11. Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* **185**, 345–351
12. Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* **110**, 87–93
13. Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367
14. Christensen, M., Sunde, L., Bolund, L., and Ornoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* **59**, 167–177
15. Celis, J. E., Ostergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Ornoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* **56**, 4782–4790
16. Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) In *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL
17. Ohlsson, R., Tycko, B., and Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* **14**, 435–438
18. Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121
19. Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170
20. Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpa, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* **59**, 3889–3892
21. Cui, H., Hedborg, F., He, L., Nordenskjold, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* **57**, 4469–4473
22. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* **285**, 251–254
23. Tsao, J., Yatabe, Y., Marki, I. D., Hajyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* **29**, 26–32
24. Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10632–10636
25. Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
26. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934

[CANCER RESEARCH 62, 6240–6245, November 1, 2002]

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer[1,2]

Elizabeth Hyman,[3] Päivikki Kauraniemi,[3] Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi[4]

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E. H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E. H., P. K., S. H., M. W., S. M., E. R., M. R., A. E., O. K., A. K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P. K., A. K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S. H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G. S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O. M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1–6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over 20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH[5] (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (*a*) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (*b*) identify and characterize those genes whose mRNA expres-

Fig. 1. Impact of gene copy number on global gene expression levels. *A*, percentage of over- and underexpressed genes (*Y axis*) according to copy number ratios (*X axis*). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). *B*, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7.

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

[2] Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org).

[3] Contributed equally to this work.

[4] To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkkeilijankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

[5] The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription-PCR.

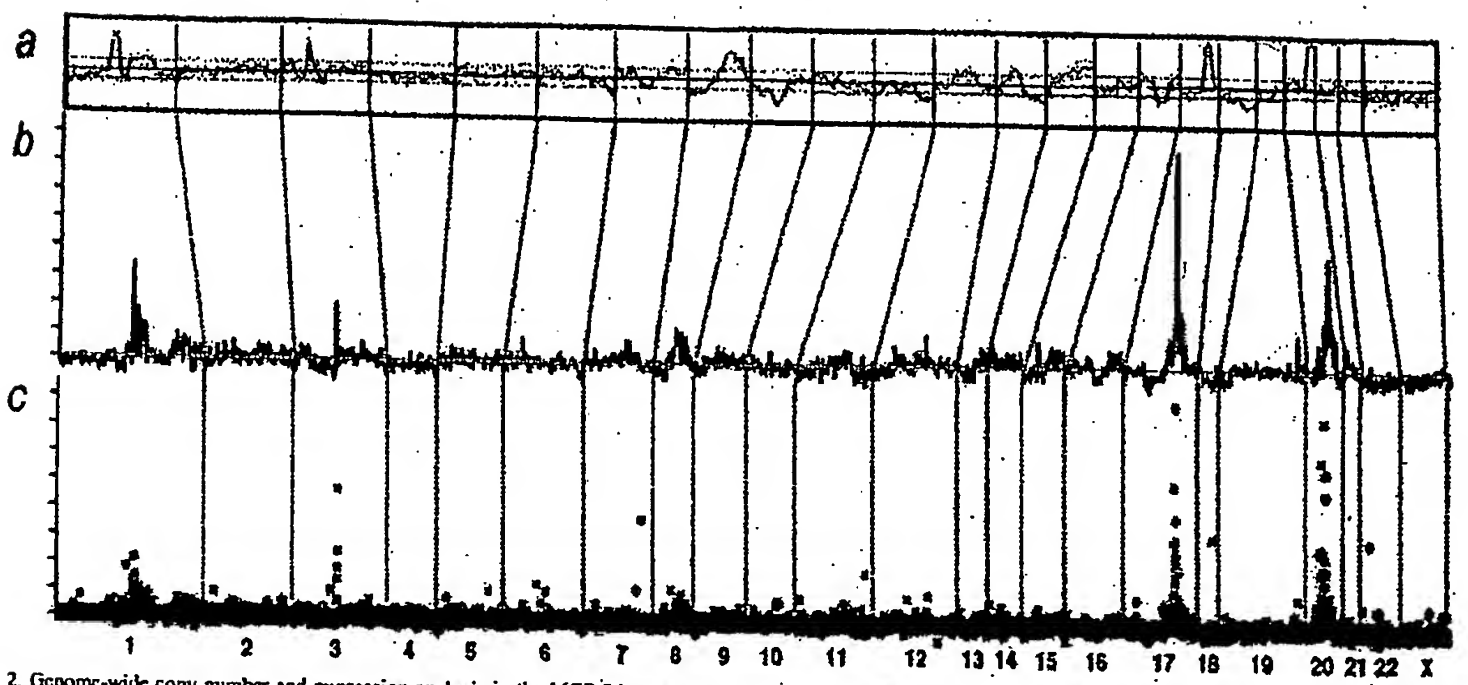


Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (*blue line*) across the entire genome from 1p telomere to Xq telomere is shown along with ±1 SD (*orange lines*). The *black horizontal line* indicates a ratio of 1.0; *red line*, a ratio of 0.8; and *green line*, a ratio of 1.2. *B–C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. *Red horizontal line*, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The *bright red dots* indicate the upper 2%, and *dark red dots*, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); *bright green dots* indicate the lowest 2%, and *dark green dots*, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with *black crosses*. The chromosome numbers are shown at the *bottom* of the figure, and chromosome boundaries are indicated with a *dashed line*.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μg of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *Alu*I and *Rsa*I (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μg of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μg of reference RNA were labeled with Cy3-dUTP and 3.5 μg of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units) were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g, for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1}, σ_{g1} and m_{g0}, σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α. A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.[6] A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.[7] The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

[6] Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.
[7] Internet address: www.genome.ucsc.edu.

Table 1 *Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray*

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1–7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11–8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3–8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22–q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12–q21.2	39.79	42.80	3.0
17q21.32–q21.33	52.47	55.80	3.3
17q22–q23.3	63.81	69.70	5.9
17q23.3–q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12–q13.13	46.45	49.43	3.0
20q13.2–q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5′-GAGCAGAGGGACTCGGACTT-3′ and 5′-GCGTCAGGTAGCGATTGTAG-3′.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (*i.e.*, belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1*A*). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1*B*). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. *A*). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal CGH were validated, with 1q21, 17q12–q21.2, 17q22–q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65–39.25 Mb), and 17q21.3 (52.47–55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2*C* shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22–q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11–p12 (Fig. 3*A*). In BT-474, the two known amplicons at 17q12 and 17q22–q23 contained numerous highly overexpressed genes (Fig. 3*B*). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3*B*, *inset*) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

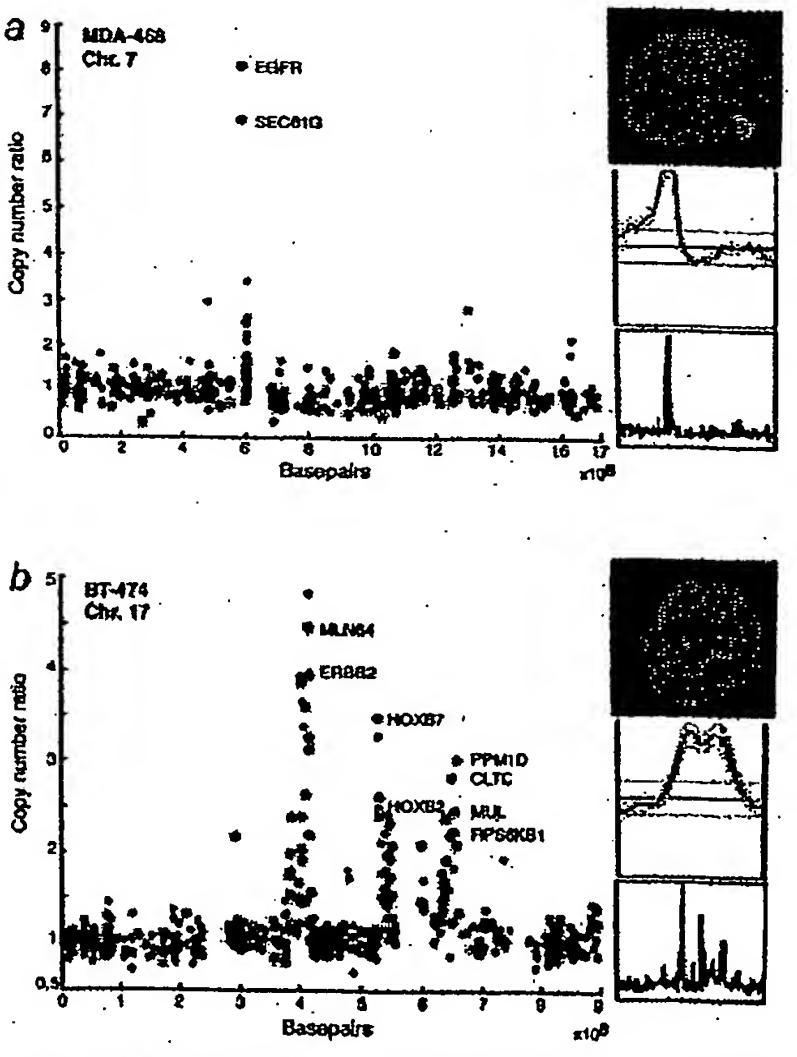


Fig. 3. Annotation of gene expression data on CGH microarray profiles. *A*, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (*red dots*) and include the *EGFR* oncogene. *B*, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (*red*) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2*C*. *Insets* show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (*red*) and chromosome 7 centromere probe (*green*) to MDA-468 (*A*) and *HOXB7*-specific probe (*red*) and chromosome 17 centromere (*green*) to BT-474 cells (*B*).

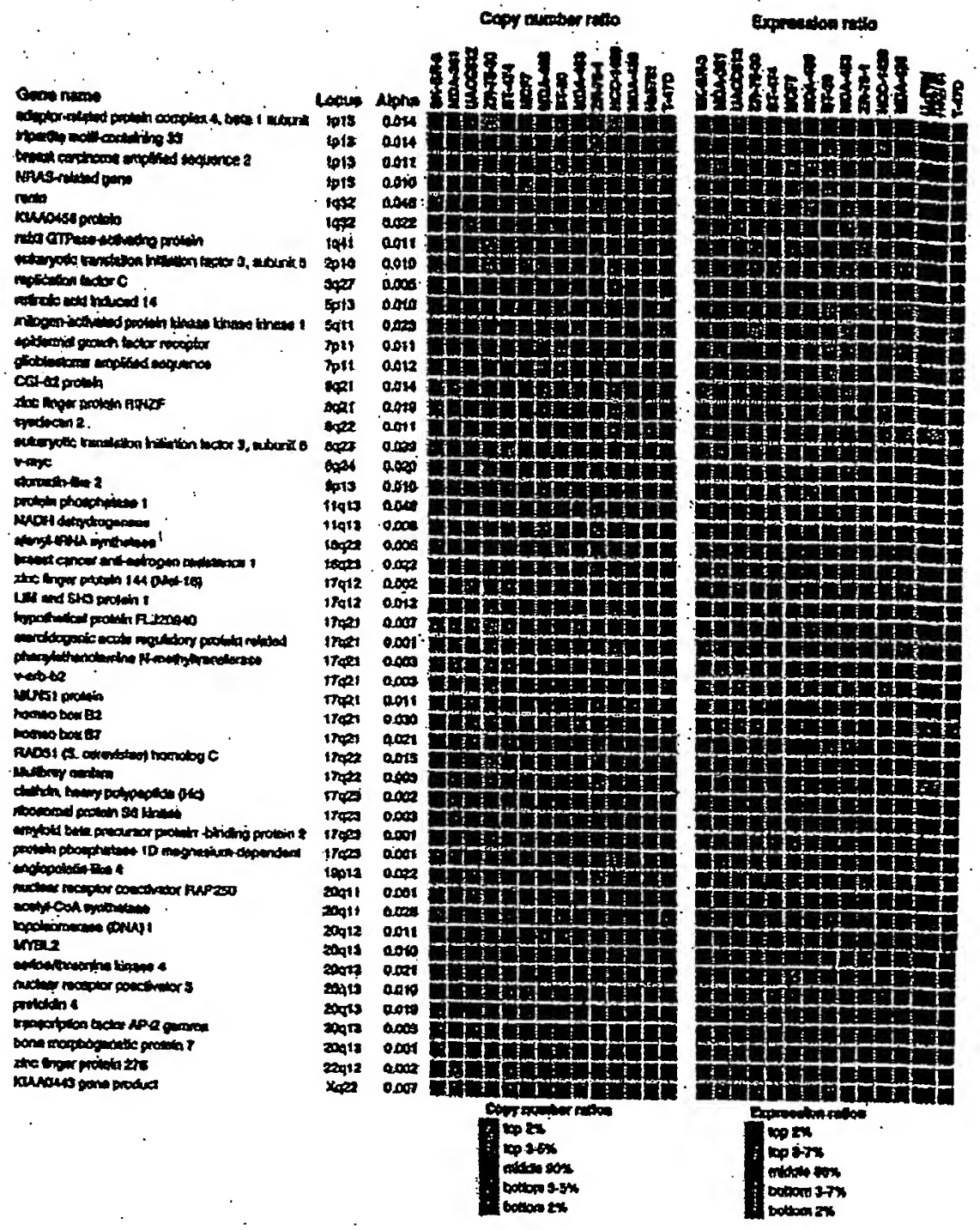


Fig. 4. List of 50 genes with a statistically significant correlation (α value <0.05) between gene copy number and gene expression. Name, chromosomal location, and the α value for each gene are indicated. The genes have been ordered according to their position in the genome. The color maps on the *right* illustrate the copy number and expression ratio patterns in the 14 cell lines. The key to the color code is shown at the *bottom* of the graph. *Gray squares*, missing values. The complete list of 270 genes is shown in supplemental Fig. B.

amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ($P = 0.001$).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,[8] 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

[8] Internet address: http://www.geneontology.org/.

DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH[9] (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

[9] Internet address: http://www.ncbi.nlm.nih.gov/entrez.

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein s6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (*a*) evidence of a prominent global influence of copy number changes on gene expression levels; (*b*) a high-resolution map of 24 independent amplicons in breast cancer; and (*c*) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

1. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.
2. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), *403:* 503–511, 2000.
3. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., *et al.* Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature (Lond.), *406:* 536–540, 2000.
4. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.
5. Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. Nature (Lond.), *412:* 822–826, 2001.
6. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA, *98:* 10869–10874, 2001.
7. Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. Semin. Cancer Biol., *9:* 125–138, 1999.
8. Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. J. Clin. Oncol., *19:* 32–40, 2001.
9. Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., *et al.* DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am. J. Pathol., *152:* 1107–1123, 1998.
10. Knuutila S., Autio K., and Aalto Y. Online access to CGH data of DNA sequence copy number changes. Am. J. Pathol., *157:* 689, 2000.
11. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.
12. Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res., *6:* 639–645, 1996.
13. Moussas, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. *In:* F. J. Livesey and S. P. Hunt (eds.), Functional Genomics, pp. 113–137. Oxford: Oxford University Press, 2000.
14. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet., *23:* 41–46, 1999.
15. Monni, O., Bärlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc. Natl. Acad. Sci. USA, *98:* 5711–5716, 2001.
16. Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics, *2:* 364–374, 1997.
17. Bärlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., *et al.* Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J. Natl. Cancer Inst., *92:* 1252–1259, 2000.
18. Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. Cytometry, *45:* 83–86, 2001.
19. Kauraniemi, P., Bärlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res., *61:* 8235–8240, 2001.
20. Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer, *34:* 104–114, 2002.
21. Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res., *62:* 2625–2629, 2002.
22. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet., *25:* 333–337, 2000.

23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc. Natl. Acad. Sci. USA, *98:* 1124–1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linehan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res., *61:* 8143–8149, 2001.
25. Bärlund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22–q24 by CGH in breast cancer is due to high-level amplification of two separate regions. Genes Chromosomes Cancer, *20:* 372–376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. Cancer Res., *56:* 3441–3445, 1996.
27. Cillo, C., Faicila, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. Exp. Cell Res., *248:* 1–9, 1999.
28. Cillo, C., Cantile, M., Faicila, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. J. Cell. Physiol., *188:* 161–169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. Mol. Cell. Biol., *16:* 4842–4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the *HOXB7* gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. Oncogene, *16:* 3285–3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. Cancer Res., *61:* 6532–6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. Proc. Natl. Acad. Sci. USA, *98:* 4060–4065, 2001.

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack*†‡, Therese Sørlie§, Charles M. Perou¶, Christian A. Rees‖**, Stefanie S. Jeffrey††, Per E. Lonning‡‡, Robert Tibshirani§§, David Botstein¶, Anne-Lise Børresen-Dale§, and Patrick O. Brown†¶¶

Departments of *Pathology, ¶Genetics, ††Surgery, §§Health Research and Policy, and ¶¶Biochemistry, and †Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; §Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ‡‡Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ‖Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. "Test" DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The "reference" (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at http://rana.lbl.gov). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

‡To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

**Present address: Zyomyx Inc., Hayward, CA 94545.

www.pnas.org/cgi/doi/10.1073/pnas.162471999

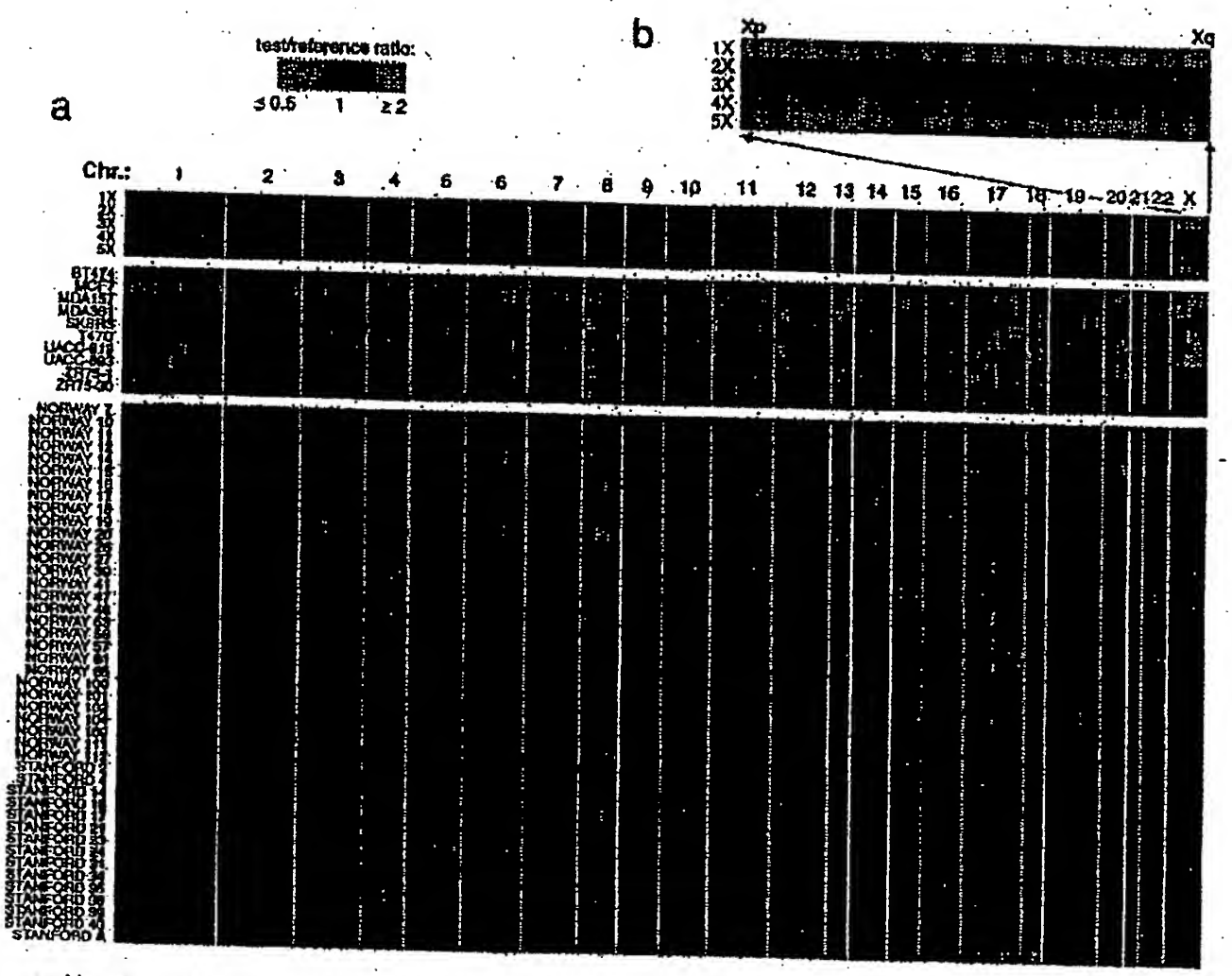


Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (*a*) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a $\log_2$-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (*b*) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (http://genome.ucsc.edu/; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1*a*; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (http://genome.ucsc.edu/) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1*b*), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1*a*), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2–4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total

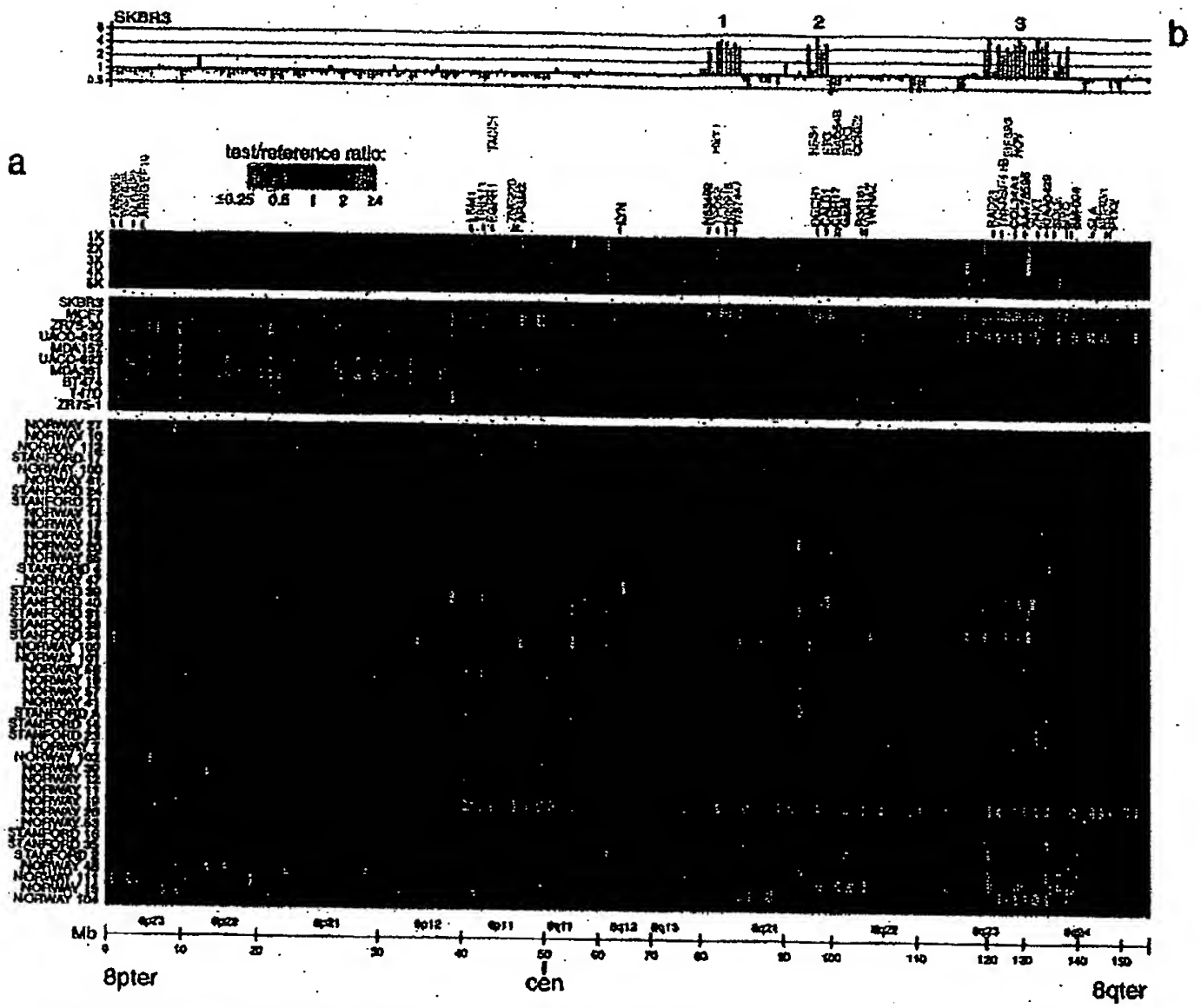


Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (*a*) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a $\log_2$ pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (*b*) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a $\log_2$ scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2*a*). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2*b*). For each of these regions we can define the boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

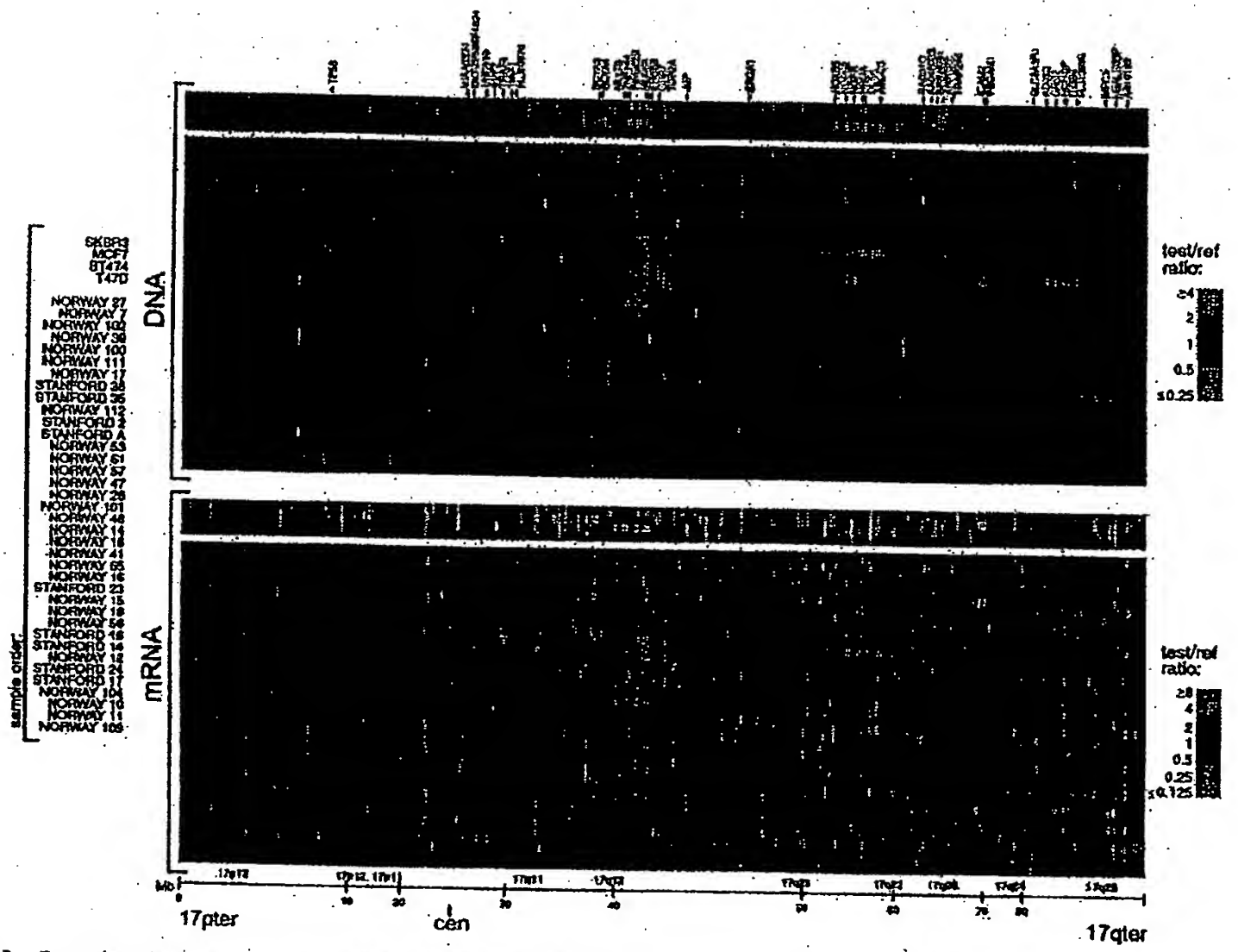


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (*Upper*) and mRNA levels (*Lower*) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (*Upper*), and the identical sample order is maintained (*Lower*). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate $\log_2$ pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4, and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4*a*). For both the breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (P values for pair-wise Student's t tests comparing adjacent classes: cell lines, 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2}; tumors, 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4}). A linear regression of the average log(DNA copy number), for each class, against average log(mRNA level) demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4*a*, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4*b*). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4*c*), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4*b*) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37

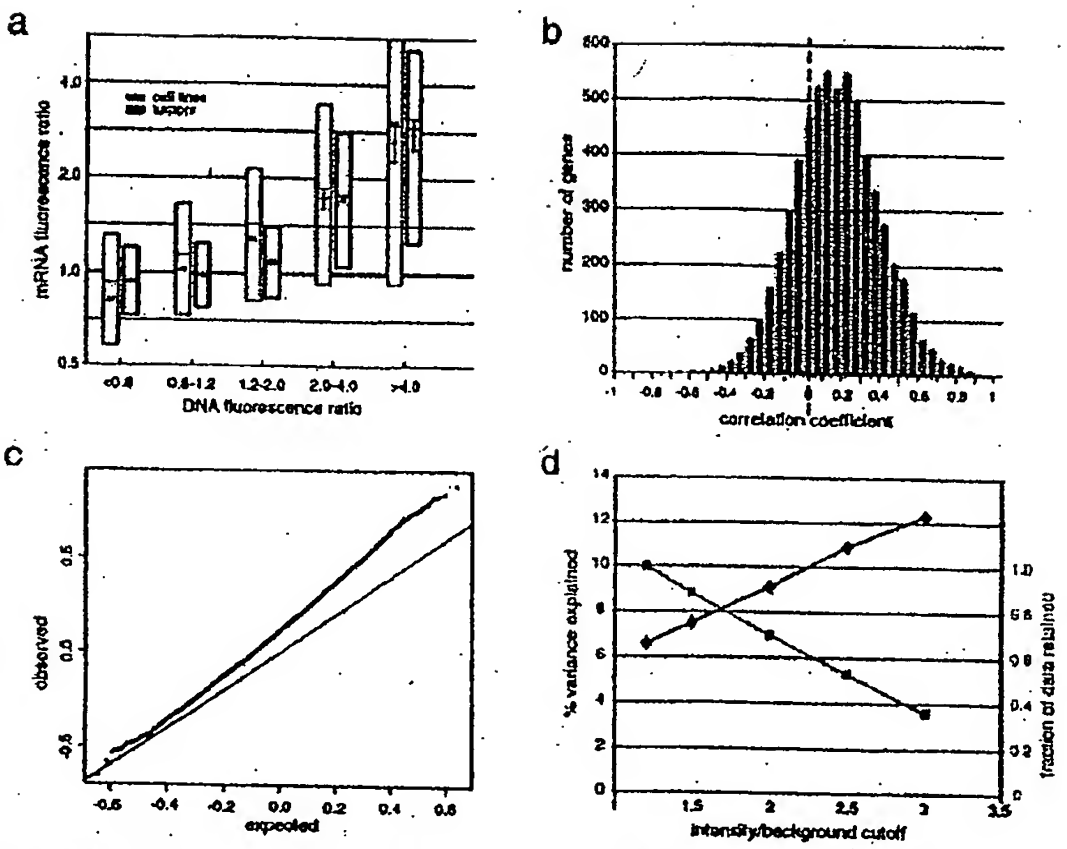


Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (*a*) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio ($\log_2$ scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; *Y*-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49}, 1×10^{-69}, 5×10^{-5}, 1×10^{-2} (cell lines), and 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4} (tumors). (*b*) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (*c*) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (*d*) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4*d*). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background >3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4*d*). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stochiometric relationships in cell metabolism and physiology (e.g., proteosome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* **258**, 818–821.
2. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2156–2160.
3. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* **21**, 177–184.
4. Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* **60**, 4519–4525.
5. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* **20**, 399–407.
6. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* **20**, 207–211.
7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* **23**, 41–46.
8. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* **406**, 747–752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874.
10. Schuler, G. D. (1997) *J. Mol. Med.* **75**, 694–698.
11. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* **409**, 860–921.
12. Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* **23**, 105–113.
13. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* **25**, 333–337.
14. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* **61**, 8143–8149.
15. Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* **62**, 1134–1138.
16. Albertson, D. G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* **25**, 144–146.
17. Li, R., Yergunian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14506–14511.
18. Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* **340**, 621–630.

BMC Gastroenterology

BioMed Central

Research article

Open Access

Cyclin A and cyclin D1 as significant prognostic markers in colorectal cancer patients

Abeer A Bahnassy*[1], Abdel-Rahman N Zekri[2], Soumaya El-Houssini[1], Amal MR El-Shehaby[3], Moustafa Raafat Mahmoud[1], Samira Abdallah[4] and Mostafa El-Serafi[5]

Address: [1]Pathology Department, National Cancer Institute, Cairo University, Cairo, Egypt, [2]Virology and Immunology Unit, Cancer Biology Department, National Cancer Institute, Cairo University, Cairo, Egypt, [3]Biochemistry Department, Kasr El-Eini School of Medicine, Cairo University, Cairo, Egypt, [4]Pathologyy Department, Kasr El-Eini School of Medicine, Cairo University, Cairo, Egypt and [5]Medical Oncology Department, National Cancer Institute, Cairo University, Cairo, Egypt

Email: Abeer A Bahnassy* - chaya2000@hotmail.com; Abdel-Rahman N Zekri - ncizakri@starnet.com.eg; Soumaya El-Houssini - chaya2000@hotmail.com; Amal MR El-Shehaby - chaya2000@hotmail.com; Moustafa Raafat Mahmoud - ncizakri@starnet.com.eg; Samira Abdallah - chaya2000@hotmail.com; Mostafa El-Serafi - melserafi@starnet.com.eg

* Corresponding author

Published: 23 September 2004

BMC Gastroenterology 2004, 4:22 doi:10.1186/1471-230X-4-22

Received: 25 April 2004
Accepted: 23 September 2004

This article is available from: http://www.biomedcentral.com/1471-230X/4/22

Abstract

Background: Colorectal cancer is a common cancer all over the world. Aberrations in the cell cycle checkpoints have been shown to be of prognostic significance in colorectal cancer.

Methods: The expression of *cyclin D1*, *cyclin A*, *histone H3* and *Ki-67* was examined in 60 colorectal cancer cases for co-regulation and impact on overall survival using immunohistochemistry, southern blot and in situ hybridization techniques. Immunoreactivity was evaluated semi quantitatively by determining the staining index of the studied proteins.

Results: There was a significant correlation between *cyclin D1* gene amplification and protein overexpression (concordance = 63.6%) and between *Ki-67* and the other studied proteins. The staining index for *Ki-67, cyclin A and D1* was higher in large, poorly differentiated tumors. The staining index of *cyclin D1* was significantly higher in cases with deeply invasive tumors and nodal metastasis. Overexpression of *cyclin A* and *D1* and amplification of *cyclin D1* were associated with reduced overall survival. Multivariate analysis shows that *cyclin D1 and A* are two independent prognostic factors in colorectal cancer patients.

Conclusions: Loss of cell cycle checkpoints control is common in colorectal cancer. *Cyclin A and D1* are superior independent indicators of poor prognosis in colorectal cancer patients. Therefore, they may help in predicting the clinical outcome of those patients on an individual basis and could be considered important therapeutic targets.

Background

Colorectal cancer (CRC) is the third most common cancer in Western countries [1]. In Egypt, CRC has unique characteristics that differ from that reported in other countries of the western society. It was estimated that 35.6% of the Egyptian CRC cases are below 40 years of age and patients

usually present with advanced stage, high grade tumors that carry more mutations [2]. This uniquely high proportion of early-onset CRC, the early and continuous exposure to hazardous environmental agents, the different mutational spectrum and the prevalent consanguinity in Egypt justify further studies [3]. It was proved that most cancers result from accumulation of genetic alterations involving certain groups of genes, the majority of which are cell cycle regulators that either stimulate or inhibit cell cycle progression [1]. Cell proliferation allows orderly progression through the cell cycle, which is governed by a number of proteins including *cyclins* and *cyclin* dependent kinases [4,5]. The *cyclins* belong to a superfamily of genes whose products complex with various *cyclin*-dependent kinases (*cdks*) to regulate transitions through key checkpoints of the cell cycle [6]. Abnormalities of several *cyclins* have been reported in different tumor types, implicating, in particular, *cyclin A*, *cyclin E* and *cyclin D* [6,7].

Cyclin D1 is a G1 *cyclin* that regulates the transition from G1 to S phase since its peak level and maximum activity are reached during the G1 phase of the cell cycle. Whereas *cyclin A* is regarded a regulator of the transition to mitosis since it reaches its maximum level during the S and G2 phases [8]. The mechanisms likely to activate the oncogenic properties of the *cyclins* include chromosomal translocations, gene amplification and aberrant protein overexpression [7,9].

Several studies have shown that, *histone H3* mRNA expression can be used to identify the S phase fraction (SPF) through the in situ hybridization (ISH) technique [10,11]. The level of *histone H3* mRNA reaches its peak during the S phase and then drops rapidly at the G2 phase [12].

In face of the increasing incidence of CRC and its peculiar pattern in the Egyptian population, the present study was conducted to assess the role of *Ki-67* (pan-cell cycle marker), *cyclin D1* (G1 phase marker), *histone H3* mRNA (S phase marker), *cyclin A* (S to G2 phase marker) in CRC. The expression level of these markers was correlated to the clinicopathologic features and the overall survival of patients.

Table 1: Clinicopathological features of patients in relation to the staining index (SI) of *Ki-67*, *cyclin D1*, *cyclin A*, *histone H3*

		SI (mean + SD)			
Variables	*No. of cases*	*Ki-67*	*Cyclin D1*	*Cyclin A*	*Histone H3*
Sex					
Male	36	18.0 ± 6.4	6.7 ± 4.3	12.7 ± 5.7	10.7 ± 5.3
Female	24	20.1 ± 5.8	8.8 ± 8.4	10.0 ± 6.0	10.7 ± 5.4
Age (years)					
≥50	41	11.7 ± 6.0*	5.6 ± 5.2	10.0 ± 5.3	6.0 ± 5.0*
<50	19	23.8 ± 5.6	7.7 ± 6.8	13.6 ± 5.7	22.0 ± 5.2
Tumor size (cm)					
<5.0	33	12.2 ± 6.3*	5.3 ± 3.8*	11.5 ± 6.1*	10.3 ± 4.9*
≥5.0	27	30.1 ± 6.2	22.8 ± 7.2	28.6 ± 5.6	24.0 ± 5.6
Histology					
Normal	20	3.5 ± 2.0*	0.6 ± 0.2*	2.3 ± 1.1*	2.2 ± 0.9
Carcinoma	60	30.3 ± 6.2	24.9 ± 6.3	27.2 ± 5.8	10.7 ± 5.3
GI	15	11.7 ± 6.2	6.6 ± 4.0	10.0 ± 5.4	11.4 ± 4.9
GII	21	11.8 ± 5.6	8.9 ± 3.6	12.3 ± 6.5	7.8 ± 5.4
GIII	24	30.0 ± 4.3	22.0 ± 8.1	27.0 ± 4.9	11.5 ± 5.4
Lymph node					
Negative	33	19.5 ± 7.0	5.4 ± 5.3*	11.9 ± 6.5	12.3 ± 5.5
Positive	27	21.3 ± 4.9	20.6 ± 6.9	12.5 ± 5.0	14.2 ± 5.0
Depth of invasion					
m, sm	17	20.7 ± 6.7	3.1 ± 3.1*	11.9 ± 7.2	10.4 ± 5.1
beyond sm	43	21.9 ± 6.2	12.4 ± 6.5	12.2 ± 5.6	10.7 ± 5.4
Stage					
I	6	20.6 ± 6.7	5.7 ± 6.9	24.2 ± 6.9	11.1 ± 5.3
II	27	20.8 ± 6.9	5.3 ± 4.3	24.6 ± 6.0	10.4 ± 5.7
III	12	22.0 ± 5.4	7.7 ± 6.0	27.1 ± 5.2	10.4 ± 4.9
IV	15	24.7 ± 6.1	11.3 ± 9.6	27.5 ± 5.5	12.3 ± 6.2

* p. value < 0.05 (significant)

Methods

Tissue samples

Paraffin-embedded tumor tissues were obtained from 60 CRC patients (47 colon and 13 rectal carcinomas) that were diagnosed and treated at the National Cancer Institute, Cairo, Egypt during the period from January, 1997 to June, 2002. Clinicopathological data of the studied cases are illustrated in table 1. None of the patients received any chemotherapy or irradiation prior to surgery. Histological diagnosis of all cases was done by 2 independent pathologists according to the WHO Histological Classification. Tumors were staged according to the TNM staging system [13]. The depth of tumor invasion was classified as invasion of the mucosa including muscularis mucosa (m), invasion of the submucosa (sm), or invasion beyond the submucosa [8]. Normal colonic tissues were obtained from autopsy specimens (n = 20) and were used as a control. The actual survival rate of the patients was calculated from the date of resection to the date of death.

Immunohistochemistry

Four micron sections of each normal and tumor specimen were cut onto positive-charged slides; air dried overnight, de-paraffinized in xylene, hydrated through a series of graded alcohol and washed in distilled water and 0.01 PBS (pH 7.4). Slides were then processed for IHC as described by Handa et al. [8]. using the following antibodies: *Ki-67* (MIB-1, Dako), *cyclin A* (6E6; Novocastra, Newcastle-Upon-Tyne, UK) and *cyclin D1* (DCS-6, Dako). A case of invasive breast cancer was used as a positive control for *Ki-67* and *cyclin A* whereas a case of mantle cell lymphoma was used as a control for *cyclin D1*. Negative controls were obtained by replacing the primary antibody by non-immunized rabbit or mouse serum.

Brown nuclear staining was regarded as a positive result for all studied markers. The proportion of positively-stained cells and the intensity of staining were scored in tumor and normal colorectal mucosal sections at medium power (×200). The degree of positive tumor staining (percentage of positive tumor cells in the examined section) was scored from 1–6 and the staining intensity was scored from 0–6 according to the pattern of staining in the examined section. Staining index (SI) was calculated by multiplying the cellularity and staining scores as described by King et al. [14].

In situ hybridization

All tumor samples and 5 normal controls were assessed for *histone H3* mRNA by ISH using the commercially available 550 base fluorescein-labeled DNA probe (Dako, Carpinteria, CA) as described by Nagao et al., 1996. This probe hybridizes to the whole mRNA transcript of the human *histoneH3* gene including the5' and 3' un-translated regions. Scoring of *histone H3* mRNA was performed as for immunohistochemistry, however, hybridization signals were detected in the cytoplasm.

Molecular detection of cyclin D1 gene amplification

High molecular weight DNA was extracted from paraffin-embedded tissues of the tumor and normal colorectal mucosal samples as previously described [15]. The proportion of neoplastic and normal cells was determined in each tumor sample by examining hematoxylin and eosin-stained slides obtained from the edge of the specimen used for DNA extraction. Tumor samples were evaluated for amplification of *cyclin D1* if more than 75% of the examined sections were formed of neoplastic cells. Accordingly, 50 cases were eligible for the analysis. Ten micrograms of the extracted DNA was digested with *Eco*R1. DNA from selected cases was also digested with *Bgl*II and *Hind*III. Samples were separated on 0.8% agarose gels and transferred to Hybond-N membranes (Amersham Int., Amersham, UK). The membranes were hybridized with 50% formamide, 5 × SSC, 5 × Denhardt's, 500 μg/ml denatured salmon sperm DNA, 10% dextran sulphate and 10^6 cpm/ml of ^{32}P-labeled *PRAD-1* probe for 24 h. Membranes were washed with 2 × SSC, 0.1% SDS at room temperature for 30 min followed by 2 × SSC, 0.1% SDS at 60°C for 30 min and 0.1 × SSC, 0.1% SDS at 60°C for 1 h. Filters were autoradiographed using an intensifying screen at -70°C for 24–72 h. After being stripped free of the *PRAD-1* probe, the same blots were hybridized with ^{32}P-labeled *B-actin* probe to normalize against possible variations in the loading or transfer of DNA. The autoradiograms were analyzed using a densitometer. Intensities of *PRAD-1/cyclin D1* were normalized to the *β-actin* control bands. The degree of amplification was calculated from these normalized values. Amplification was considered when the signal of the tumor band was ≥2-fold the value of the matched normal mucosa [16].

Statistical analysis

The Mann-Whitney non-parametric test was used to compare the SIs of pairs of subjects whereas the Kruskal-wallis was used for categorial data. Correlation between indices was performed using a simple linear regression test. The Kaplan-Meier method was used to create survival curves which were analyzed by the log-rank test. The impact of different variables on survival was determined using the Cox proportional hazards model. *p*. values less than 0.05 were considered significant.

Results

The results of IHC are illustrated in figures 1 and 2. In general, the staining index (SIs) of all studied markers was higher in carcinomas than in normal colonic mucosal samples ($p = 0.0001$). Normal colorectal mucosa revealed positive imunostaining for *Ki-67* in the lower half of the crypts only. A heterogeneous staining pattern was

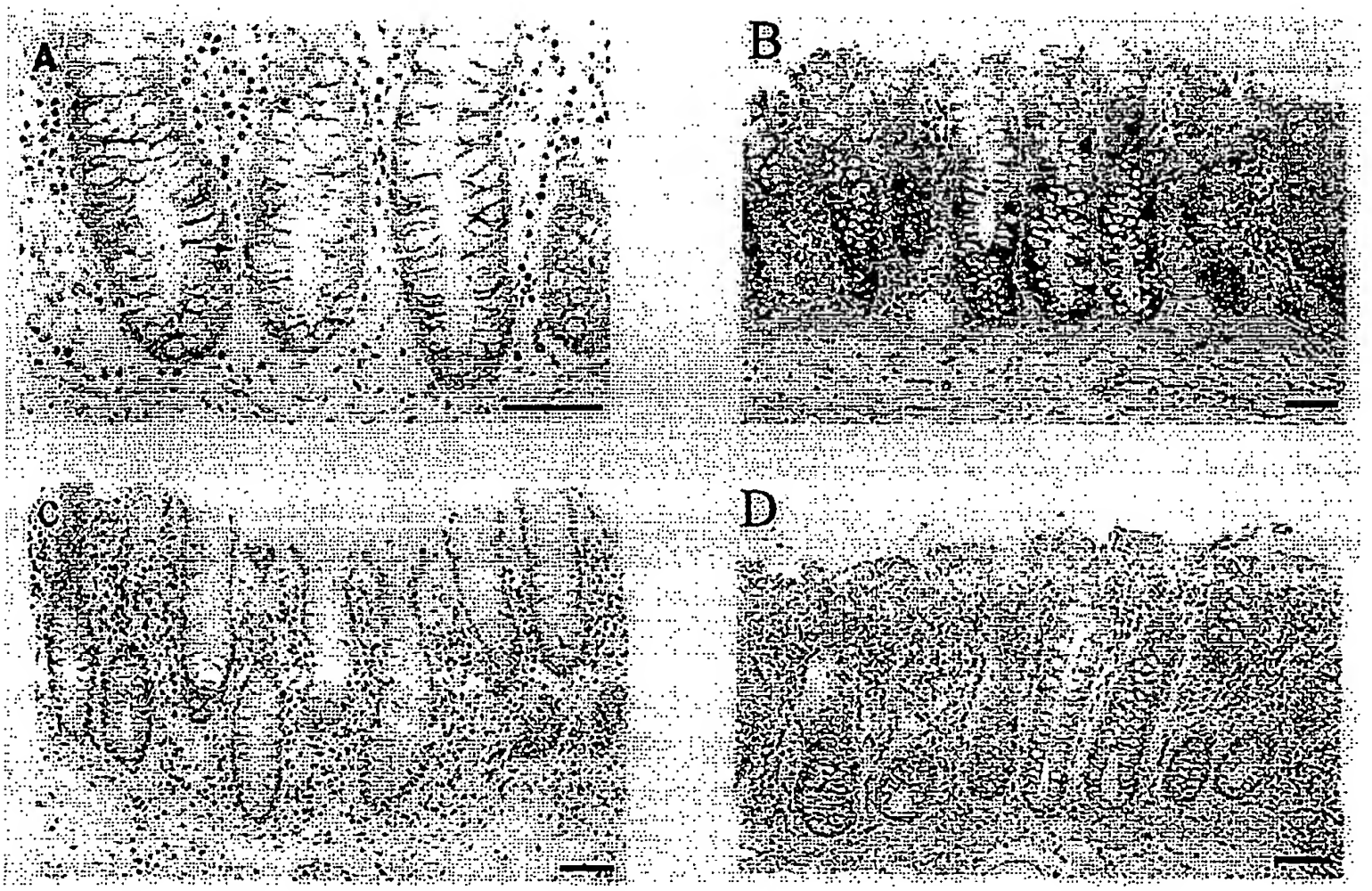


Figure 1
Normal colonic mucosa showing positive nuclear immunostaining for: (a) *cyclin D1*, (b) ISH of *histone H3* mRNA, (c) *Ki-67* and (d) *cyclin A*

detected in the neoplastic cells of well and moderately-differentiated adenocarcinomas whereas a diffuse homogeneous staining pattern was detected in poorly-differentiated carcinomas. The SI ranged from 10–40.2 (mean: 24.6 ± 6.5).

Immunostaining for *cyclin D1* was predominantly nuclear but cytoplasmic staining was detected in some cases. However, unless a nuclear staining was also detected, cases with cytoplasmic staining were considered negative. Normal colorectal mucosal samples were almost negative for *cyclin D1* whereas 41 out of the 60 (68.3%) CRC cases were positive. Marked heterogeneity was observed in well- and moderately-differentiated adenocarcinomas even within the same tumor. Poorly-differentiated carcinomas revealed a diffuse staining pattern with more darkly-stained nuclei. The SI ranged from 0.5–28.6 (mean: 9.3 ± 4.2).

Positive nuclear staining for *cyclin A* was detected in 80% (48/60) of CRC cases and in all non-neoplastic control samples. Positively-stained nuclei were confined to the lower half of the crypts in normal colonic mucosa and diffusely-dispersed in carcinomas. The SI ranged from 3.3–30.2 (mean: 15.1 ± 6.6).

Histone H3 mRNA was intensely expressed in the cytoplasm of all examined samples either neoplastic or non-neoplastic. The distribution of *histone H3* mRNA was similar to that of *cyclin A* and *Ki-67* however, the proportion of *histone H3* mRNA positive cells was less than that of *Ki-67*. The SI ranged from 1.8–24.2 (mean: 12.4 ± 5.3).

The *PRAD-1* probe detected 3 *Eco*RI fragments of 4.0, 2.2 and 2.0 and 1 *Bgl*II fragment of 15 Kb. *PRAD-1/cyclin D1* gene amplification was detected in 22/50 (44%) cases analyzed. The degree of amplification was heterogeneous

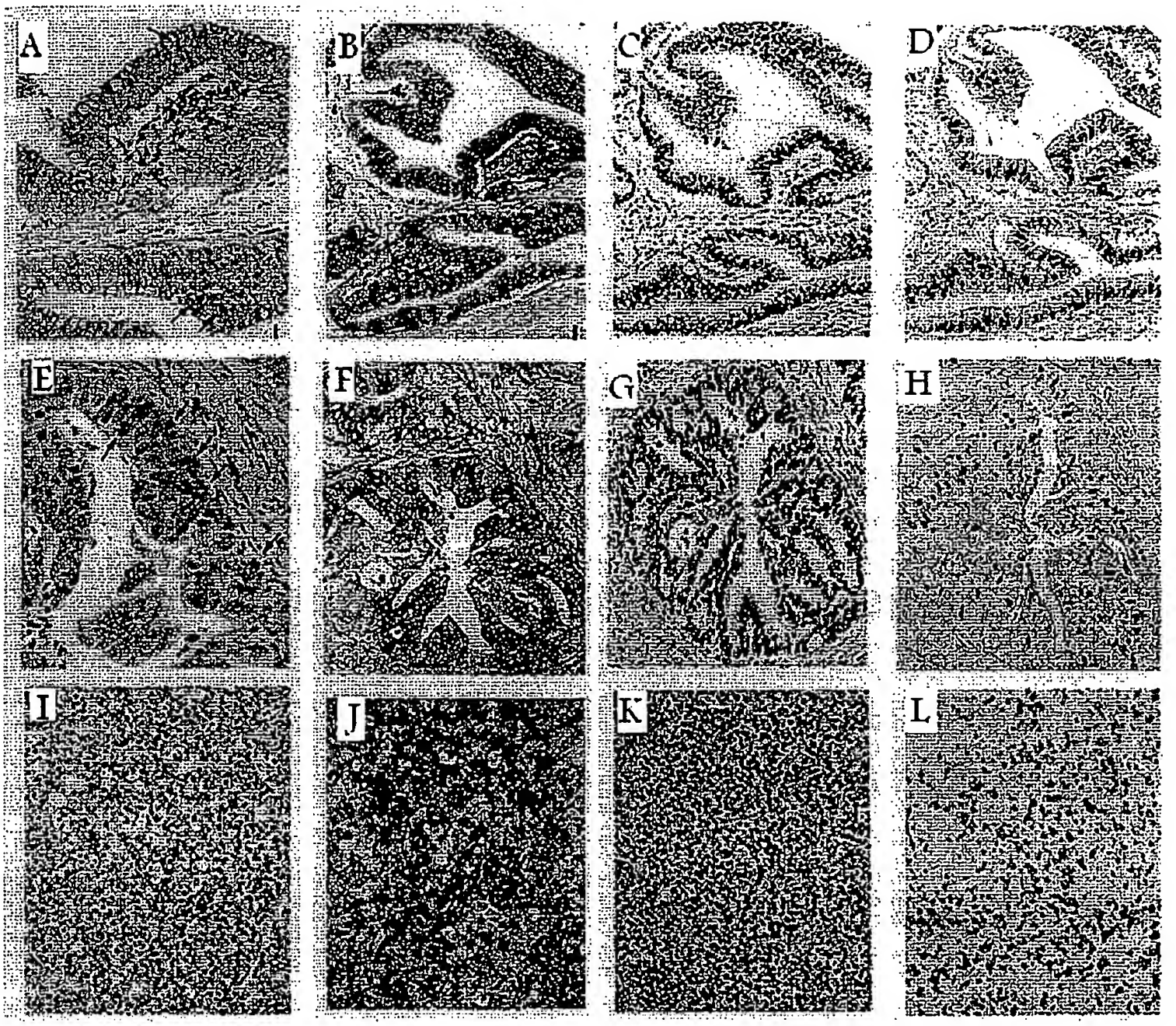


Figure 2
A case of well differentiated adenocarcinoma with positive immunostaining for: (a) *cyclin D1*, (b) *histone H3* mRNA, (c) *Ki-67*, and (d) *cyclin A*. Another case of moderately differentiated denocarcinoma with positive immunostaining for: (e) cyclin D1, (f) *histone H3* mRNA, (g) *Ki-67*, and (h) *cyclin A*. A case of poorly differentiated adenocarcinoma with diffuse staining for: (i) *cyclin D1*, (j) ISH of *histone H3* mRNA, (k) *Ki-67* and (l) *cyclin A*.

with 2–10 fold increase when compared to normal mucosal samples (Figure 3). Amplification was confirmed by other restriction enzymes.

Correlations

There was a significant correlation between *cyclin D1* gene amplification and protein overexpression. Out of the 22 cases that showed amplification 14 showed protein overexpression (concordance = 63.6%).

Linear regression analysis of SIs revealed a significant correlation between *Ki-67* and *cyclin D1*, *cyclin A*, *histone H3* as well as between the SIs of *cyclin A* and *histone H3* (p = 0.008, 0.0001, and 0.0001 respectively) (Figure 4). There was a significant relationship between the SI of both *Ki-67*

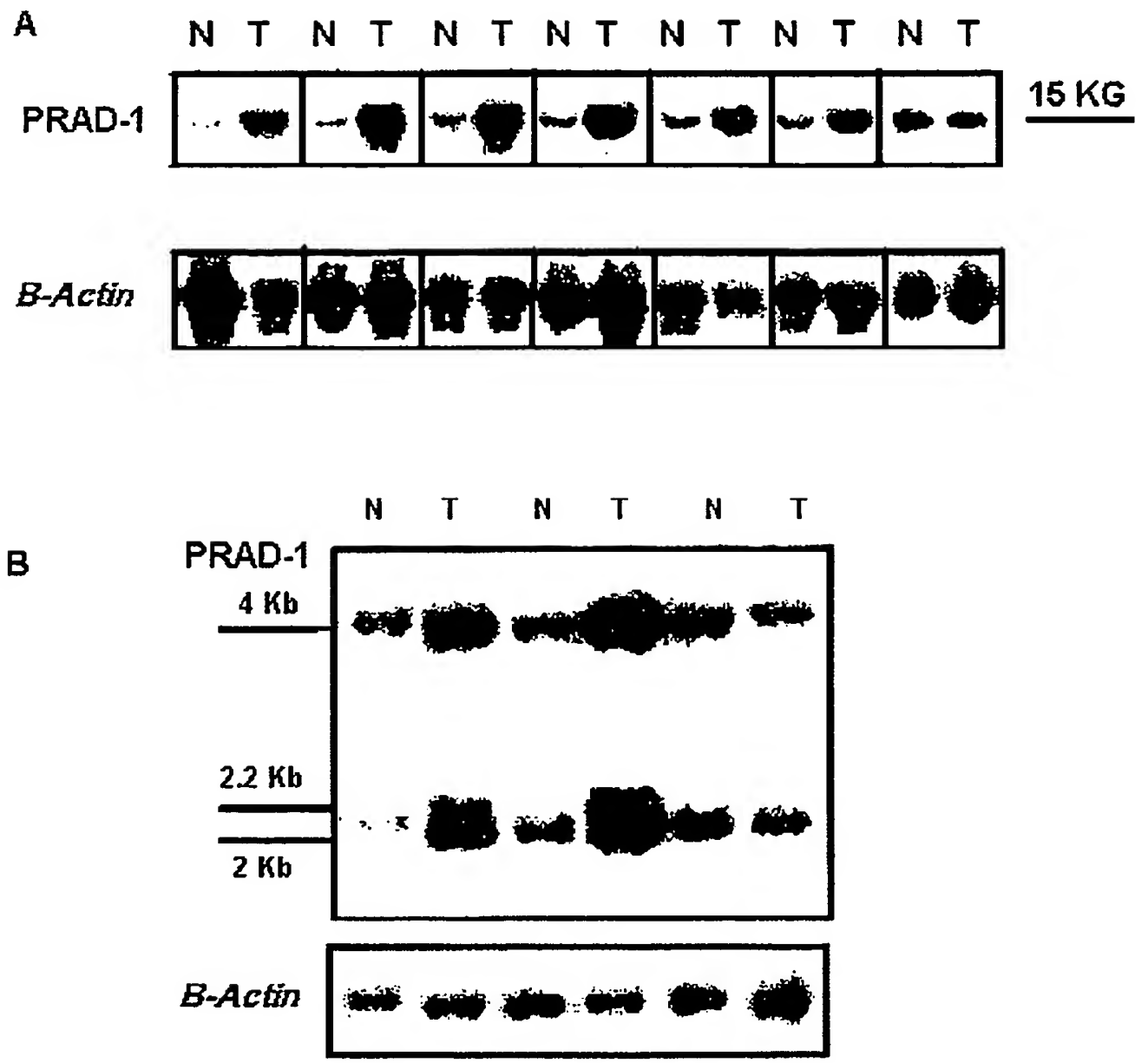


Figure 3
A: Southern blot analysis of normal mucosa (N) and their seven corresponding cases of colonic adenocarcinomas (T1–T7), cases No. 1, 2, 4, and 5 are poorly differentiated whereas cases No. 3, 6, and 7 are moderately differentiated. Genomic DNA was digested with *Bgl*II, fractionated by electrophoresis in agarose gel, transferred onto membranes and hybridized with *PRAD1* and *β-actin*. Tumors number 1–6 (Lanes 1–6) show different degrees of *PRAD1/cyclin D1* amplification, tumor number 7 (lane 7) was not amplified. **B:** Southern blot analysis of 3 cases of adenocarcinomas (T) and matched normal colonic mucosa (N). Genomic DNA was digested with *Eco*RI, fractionated by electrophoresis in agarose gel, transferred onto membranes and hybridized with *PRAD1* and *β-actin* probes for loading control. The identification of the 3 tumors is the same as in Fig. 3A with amplification of *PRAD1/cyclin D1* in tumors number 4, 5 (Lanes 1, 2) but not 7 (Lane 3).

and *cyclin A* and the degree of differentiation of tumors as well as the size of the tumor ($p < 0.001$ and $p < 0.01$ respectively). In addition, SI of *Ki-67* and *histone H3* were higher in patients <50 years than in those ≥50 years ($p < 0.05$) (table 1).

In addition table 2 shows a significant relationship between high *cyclin D1* SI and large, poorly-differentiated tumors, carcinomas with positive lymph node metastasis and deeply-invasive carcinomas ($p < 0.05$, $p < 0.001$, $p < 0.05$ and $p < 0.05$ respectively). Whereas *cyclin D1* gene amplification was significantly associated with an advanced disease stage since amplification was detected in

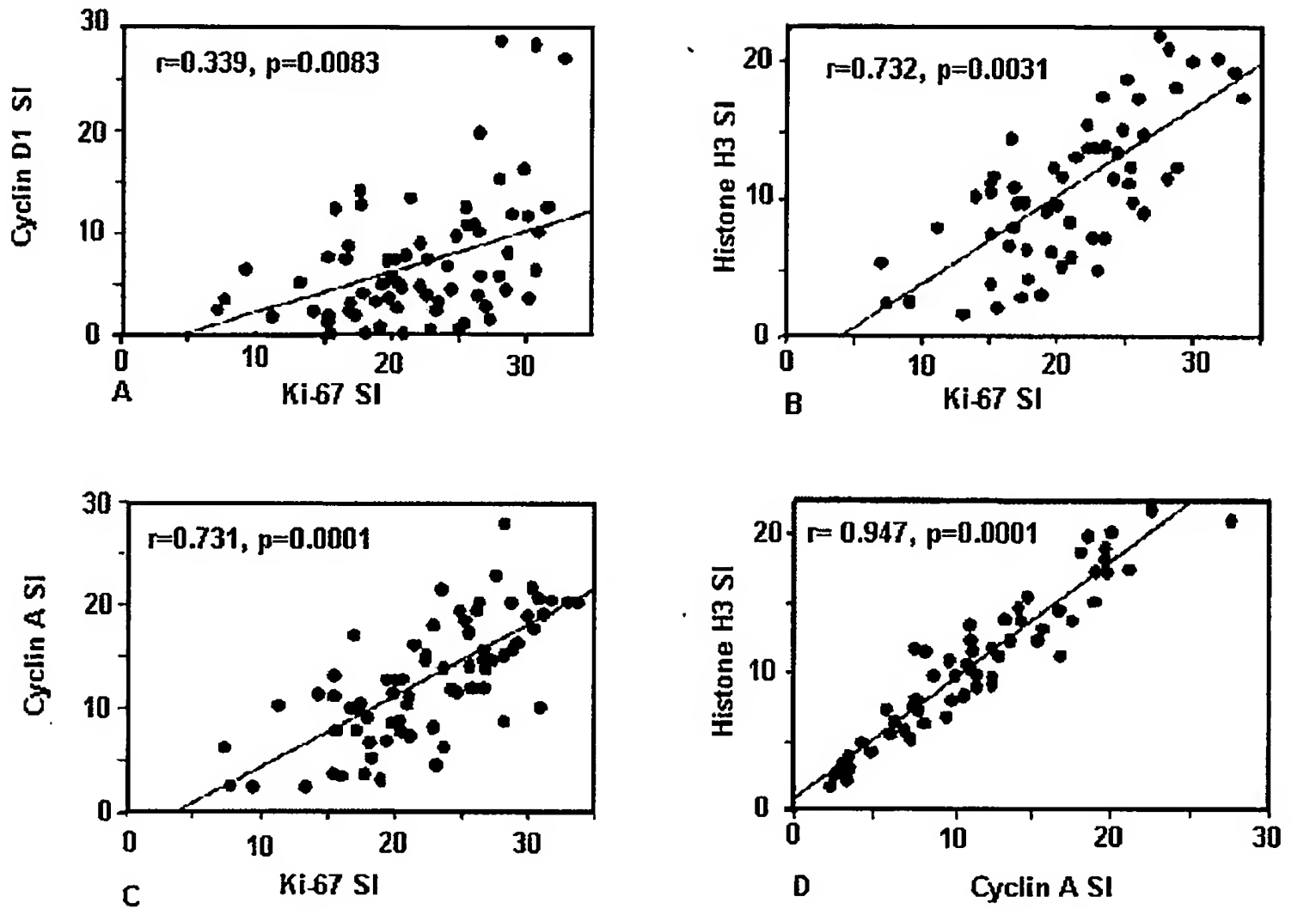


Figure 4
Correlation between the staining intensity of (a) *Ki-67 vs. cyclin D1*, (b) *Ki-67 vs. histone H3*, (c) *Ki-67 vs. cyclin A* and (d) *cyclin A vs. histone H3* mRNA expression.

10/15 (66.7%) of stage IV tumors compared to 12/45 (26.7%) of stage I-III tumors ($p = 0.002$). Similarly, DNA amplification was detected in 60.5% (26/43) of the carcinomas with extensive local invasion (beyond sm) but only in 23.5% (4/17) of the carcinomas with limited invasion (m, sm) ($p = 0.001$). A significant correlation was also present between *cyclin D1* gene amplification and the presence of lymph node metastasis ($p = 0.008$) as well as between the SI of *histone H3*, the size of the tumor and the patient's age ($p < 0.05$, $p < 0.001$ respectively). The SI was higher in tumors >5 cm in diameter and in patients <50 years.

Survival analysis

The mean follow-up period for all patients was 30 months (range: 1–66 months). Eighteen of 60 patients had already died by the time the study was completed. We defined the cutoff level for overexpression of each cell cycle marker at the point that showed the maximum difference of survival rate between the 2 groups separated by that point. Cox regression analysis revealed that *cyclin A* overexpression (our definition: SI ≥ 10.5), *cyclin D1* overexpression (our definition: SI ≥ 6.1), poorly differentiated histology, lymph node metastasis, TNM stage, tumor size and depth of invasion were all significant prognostic variables for survival (Table 3). The Kaplan-Meier survival curves for the subgroups of patients who are subdivided according to each marker's status are shown in Figure 5. Patient with tumors that showed *Ki-67* overexpression (our definition: SI ≥ 11.5) and *histone H3* overexpression (our definition: SI ≥ 8.2) tended to have poor prognosis but this did not reach a statistically significant level, however the overall survival was significantly lower in patient with *cyclin A* and *cyclin D1* overexpression. Cox multivari-

Table 2: The relation between *cyclin D1* overexpression vs *cyclin D1* amplification and clinicopathological prognostic markers.

Variables	*No. of cases*	*Cyclin D1 overexpression*	*Cyclin D1 Amplification*
Tumor size (cm)			
<5.0	33	5.3 ± 3.8*	13/33
≥5.0	27	22.8 ± 7.2 p *<0.05*	9/27 p *<0.236*
Histology			
GI	15	6.6 ± 4.0	7/15
GII	21	8.9 ± 3.6	8/21
GIII	24	22.0 ± 8.1 p *<0.001*	7/24 p *<0.075*
Lymph node			
Negative	33	5.4 ± 5.3*	6/33 (18.2%)
Positive	27	20.6 ± 6.9 p *<0.05*	16/27 (59.3%) p *<0.008*
Depth of invasion			
m, sm	17	3.1 ± 3.1*	4/17 (23.5%)
beyond sm	43	12.4 ± 6.5 p *<0.05*	26/43 (60.5%) p *<0.001*
Stage			
early	45	5.5 ± 10.1	12/45 (26.7%)
late	15	11.3 ± 9.6 P = *0.175*	10/15 (66.7%) p *<0.002*

Table 3: Uunivariate analysis of the relationship between survival and the tested markers

PredictiveVariables	*Median Survival*	*HR*	*CI*	*P*
Ki-67				
<11.5	36			
≥11.5	32	1.826	0.636 – 5.243	0.26
Cyclin D1				
<6.1	35			
≥6.1	18	7.246	1.007 – 45.150	0.03*
Histone H3				
<8.2	35			
≥8.2	29	4.639	0.854 – 25.196	0.07
Cyclin A				
<10.5	35			
≥10.5	15	7.820	1.017 – 60.122	0.02*
Histological grade				
Low	38			
High	10	7.331	2.696 – 19.940	0.0001*
Lymph node				
Negative	38			
Positive	15	6.826	1.973 – 23.621	0.002*
Stage				
I, II, III	38			
IV	12	6.378	1.842 – 22.083	0.001*
Tumor size (cm)				
<5.0	35			
≥5.0	13	4.835	1.386 – 16.868	0.01*
Depth of invasion				
T1, T2	36			
T3, T4	20	7.759	1.024 – 58.789	0.04*
Age (years)				
<50	38			
≥50	28	2.802	0.988 – 7.943	0.0526
Sex				
Male	38			
Female	36	0.696	00.274 – 1.766	0.4449

* p. value < 0.05 (significant)
HR: Hazard Ratio
CI: 95% confidence Interval

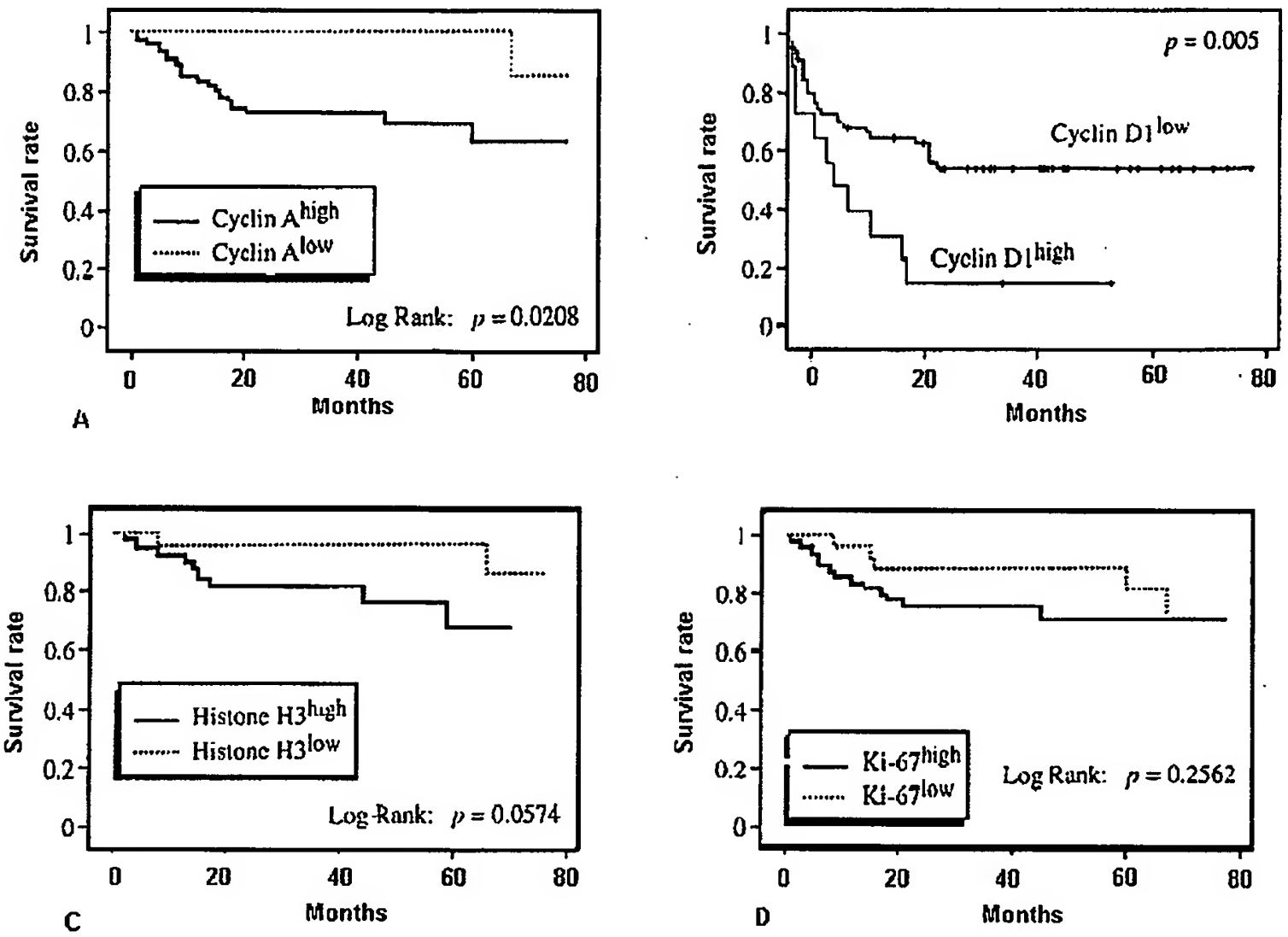


Figure 5
Kaplan-Meier survival curves for colorectal carcinoma. Overall survival is significantly lower in patients with (a) *cyclin A* and (b) *cyclin D1* overexpression. Patients with high SI for *histone H3* mRNA have poorer prognosis but this was not statistically significant (c). No significant difference was present between patients with high *Ki-67* SI and those with low *Ki-67* SI (d).

ate regression analysis revealed that lymph node metastasis, *cyclin A* and *cyclin D1* overexpression were independent negative prognostic factors after adjustment for the depth of tumor invasion, age and sex of the patient (Table 4).

Discussion

The proliferative activity of CRC cells has been investigated in several studies either by immunohistochemical determination of cell proliferation index using antibodies to some types of *cyclins* or by flowcytometric determination of the SPF of the cell cycle [8]. Although Leach et al. [17] did not find *cyclin D1* gene amplification in a panel of 47 CRC cell lines; its protein was overexpressed in about 30% of CRC cases that were included in the studies of Bartakova et al. [6] and Arber et al. [18]. In the former study [6]*cyclin D1* was aberrantly accumulated in a significant subset of human CRC cases and the cell lines derived from these cases were dependent on *cyclin* in their cell cycle progression. In the second study [18], overexpression of *cyclin D1* was detected in 30% of adenomatous polyps indicating that overexpression is a relatively early event in colon carcinogenesis which is possibly responsible for the pathological changes in the mucosa preceding neoplastic transformation. More recently, Holland et al. [19], Pasz-Walczak et al. [20] and Utsunomiya et al. [21] reported up-regulation of *cyclin D1* in 58.7%, 100% and 43% of their studied cases respectively.

Table 4: Multivariate analysis ofthe relationship between survival and thetested markers

PredictiveVariables	*HR*	*CI*	*P*
Cyclin D1	10.864	1.055 – 86.250	0.03*
(baseline < 6.1)	-	-	-
Cyclin A	13.886	1.012 – 190.579	0.0490*
(baseline < 10.5)	-	-	-
Positive Lymph node metastasis	3.921	1.057 – 14.472	0.0410*
Stage IV	3.411	1.048 – 12.083	0.03*
Depth of invasion			
T3, T4	5.408	0.449 – 65.080	0.1836
Age (years)			
≥50	1.996	0.678 – 5.878	0.2310
Sex	0.910	0.315 – 2.358	0.8453

p. value < 0.05 (significant)
HR: Hazard Ratio
CI: 95% confidence Interval

In the present study, up-regulation of *cyclin D1* was detected in 68.3% of the cases. The SI was significantly higher in carcinomas than in normal colorectal mucosa and in poorly-differentiated adenocarcinomas it was approximately twice that of other histological types. Amplification and/or overexpression of *cyclin D1* significantly correlated with deeply invasive tumors and positive lymph node metastasis. Our results in this regards are consistent with previous studies [8,22]. In 2001, Holland et al. [19]. demonstrated that deregulation of *cyclin D1* and $p21^{waf}$ proteins are important in colorectal tumorigenesis and have implications for patient prognosis. Similarly McKay et al. [23] found that *cyclin D1* was the only protein in their panel (*cyclin D1, p53, p16, Rb-1, PCNA and p27*) that correlated with improved outcome in CRC patients. However, few studies failed to detect any correlation between *cyclin D1* overexpression and the clinicopathological factors in CRC [6,18]. This controversy in results could partially be explained by the difference in the sampling of studied cases. The present study included 24 cases of poorly differentiated adenocarcinoma, which is not common in other studies of CRC in western countries. This was possible because the majority of CRC cases diagnosed in Egypt are of high histological grade [3]. The correlation between *cyclin D1* overexpression and the high histological grade was also reported in other tumor types including non-small cell lung carcinomas [24] and squamous cell carcinomas of the larynx [16]. Another possible explanation for the observed controversy in the results of different studies is the detection method used.

In the present work, overexpression of *cyclin D1* was more common than gene amplification of the *PRAD-1/cyclin D1* gene with a 63.6% concordance. This was similarly reported by Bartakova et al. [6] who mentioned that there is a subset of CRC cases in which *cyclin D1* is overexpressed without *PRAD-1/cyclin D1* gene amplification. Consistent with this hypothesis are reports of elevated *cyclin D1* mRNA levels and immunohistochemically detectable accumulation of the protein in over one third of breast cancer cases at a frequency significantly higher than that deduced from DNA amplification studies [9,25]. These data imply that mechanisms other than gene amplification can also lead to deregulation and accumulation of *cyclin D1* in solid tumors.

So far, several studies were done to reveal the prognostic significance of *cyclin D1* overexpression in various carcinomas, including CRC [22]. However, these studies yielded conflicting results which could be attributed to organ heterogeneity. In our study, patients with tumors that exhibited *cyclin D1* overexpression tended to have poor prognosis.

It was reported that, patients with *cyclin A* positive carcinomas had significantly shorter median survival times. Handa et al. [8] were able to detect *cyclin A* overexpression in 77% of their CRC cases. They also demonstrated that, *cylcin A* could be used as a prognostic factor of CRC. More recently, Habermann et al. [26] studied cases of ulcerative colitis with and without an associated adenocarcinoma for the presence of *cyclin A* overexpression. They found that, *cyclin A* overexpression was higher in cases of ulcerative colitis with adenocarcinomas than in those without adenocarcinomas. Consequently, they concluded that, *cyclin A* could be used for monitoring ulcerative colitis patients and for the early detection of an emerging carcinoma in this high risk group of patients.

In our study, *cyclin A* was detected in 80% of the patients and Cox regression analysis showed that it could be used as a prognostic marker in CRC in addition to *cyclin D1*.

It would have been useful if we assessed the expression level of *cyclin A* by another technique (DNA amplification). This would have added more information regarding the gene status on one hand and confirmed the results of IHC on the other hand. Unfortunately, this was not possible because in most of the cases included in the present work, the extracted DNA was not sufficient to study *cyclin amplification* after the assessment of *cyclin D1*.

In 1996, Nagao et al. [11] reported that *histone H3* labeling index significantly correlated with ki-67 immunostaining and was high in poorly differentiated human hepatocellular carcinoma. This was similarly reported in the present work since we found a significant correlation between the SI of *histone H3* and *Ki-67*. However, no

statistically significant correlation was found between *histone H3* SI and any of the studied clinicopathological factors.

Although *Ki-67* immunostaining reflects the proliferative activity of CRC, it has not been recognized as a significant prognostic factor in this type of tumors [27,28]. However, Suzuki at al. [29] found a significant correlation between *Ki-67* labeling index and local invasion of CRC. In the present study there was a significant relationship between the SI of *Ki-67*, tumor size and grade. However, Kaplan-Meier survival curves showed no significant difference in survival rates between patients with- and without overexpression of *Ki-67*.

Conclusions

Our results demonstrate that *cyclin D1*, *cyclin A*, *histone H3* and *Ki-67* are overexpressed in a subset of CRC, however only *cyclin D1* and *cyclin A* overexpression correlates with poor differentiation and tumor progression. This indicates the superiority of *cyclin A* and *cyclin D1* as indicators of poor prognosis compared to *Ki-67* and *histone H3* mRNA in CRC. *Cyclin A and D1* could therefore be considered significant, independent prognostic factors in CRC patients. These findings are especially important in stage II patients since 25–30% of those patients have poor prognosis in spite of being node-negative. However, the standard clinicopathologic prognostic factors can not identify this subset accurately and therefore; there is a great demand for more accurate, individually-based, biological prognostic parameters that help in detecting this high risk group of patients who can benefit from an adjuvant therapy. If the findings of the present study are confirmed in a larger study, evaluation of *cyclin A and D1* may be applicable to clinical management of CRC, allowing the identification of patients with poor prognosis.

Competing interests

The author(s) declare that they have no competing interests.

List of abbreviations

CRC – Colorectal cancer

OS – overall survival

SI – staining index

SPF – S phase fraction

ISH – in situ hybridization

m – muscularis mucosa

sm – invasion of the sub mucosa

Authors' contributions

BA and ZA-R carried out the molecular genetic studies, designed, coordinated the study and drafted the manuscript. BA and El-HS carried out all the histopathological and immunohistochemical studies. El-SA participated in molecular genetic studies and drafted the manuscript. MM coordinated the study. El-SM carried out all the patient clinical data. All authors read and approved the final manuscript

References

1. Jiang GL, Huang S: **Adenovirus expressing RIZ1 in tumor suppressor gene therapy of microsatellite unstable colorectal cancers.** *Cancer Res* 2001, 61:1796-1798.
2. Soliman AS, Bondy ML, Levin B, Hamza MR, Ismail K, Ismail S, Hammam HM, El-Hattab O, Kamal SM, Soliman AG, Dorgham LA, McPherson RS, Beasley RP: **Colorectal cancer in Egyptian patients under 40 years of age.** *Int J Cancer* 1997, 71:26-30.
3. Soliman AS, Bondy ML, Guan Y, El-Badawy S, Mokhtar N, Bayomi S, Raouf AA, Ismail S, McPherson RS, Abdel-Hakim TF, Beasley PR, Levin B, Wei Q: **Reduced expression of mismatch repair genes in colorectal cancer patients in Egypt.** *Int J Oncol* 1998, 12:1315-1319.
4. Cordon-Cardo C: **Mutations of cell cycle regulators. Biological and clinical implications for human neoplasia.** *Am J pathol* 1995, 147:545-560.
5. Hunter T, Pines J: **Cyclins and cancer. II. Cyclin D and CDK inhibitors come of age.** *Cell* 1994, 79:573-528.
6. Bartkova J, Lukas J, Strauss M, Bartek J: **The *PRAD-1/cyclin D1* oncogene product accumulates aberrantly in a subset of CRCs.** *Int J Cancer* 1994, 58:568-573.
7. Motokura T, Arnold A: **Cyclins and oncogenesis.** *Biochim Biophys Acta* 1993, 1155:63-78.
8. Handa K, Yamakawa M, Takeda H, Kimura S, Takahashi T: **Expression of the cell cycle markers in colorectal carcinoma: Superiority of cyclin A as an indicator of poor prognosis.** *Int J cancer* 1999, 84:225-233.
9. Gillett C, Fantl V, Smith R, Fisher C, Bartek J, Dickson C, Barnes D, Peters G: **Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining.** *Cancer Res* 1994, 54:1812-1817.
10. Gown AM, Jiang JJ, Matles H, Skelly M, Goodpaster T, Cass L, Reshatof M, Spaulding D, Coltrera DM: **Validation of the S-phase specificity of histone (H3) in situ hybridization in normal and malignant cells.** *J Histochem Cytochem* 1996, 44:221-226.
11. Nagao T, Ishida Y, Kondo Y: **Determination of S-phase cells by in situ hybridization for histone H3 mRNA in hepatocellular carcinoma: correlation with histological grade and other cell proliferative markers.** *Mod Pathol* 1996, 9:99-104.
12. Chou MY, Chang AL, McBride J, Donoff B, Gallagher GT, Wong DT: **A rapid method to determine proliferation patterns of normal and malignant tissues by H3 mRNA in situ hybridization.** *Am J Pathol* 1990, 136:729-733.
13. Sobin LH, Wittekind C: **TNM classification of malignant tumors.** 5th edition. John Wiley, New York; 1997.
14. King RJ, Coffer AI, Gilbert J, Lewis K, Nash R, Millis R, Raju S, Taylor RW: **Histochemical studies with a monoclonal antibody raised against a partially purified soluble estradiol receptor preparation from human myometrium.** *Cancer Res* 1985, 45:5728-5733.
15. Slebos RJ, Boerrigter L, Evers SG, Wisman P, Mooi WJ, Rodenhuis S: **A rapid and simple procedure for the routine detection of ras point mutations in formalin-fixed, paraffin-embedded tissues.** *Diag Moln Path* 1992, 1:136-141.
16. Jares P, Fernandez P, Campo E, Nadal A, Bosch F, Aiza G, Nayach I, Traserra J, Cardesa A: ***PRAD-1/cyclin D1* gene amplification correlates with messenger RNA overexpression and tumor progression in human laryngeal carcinomas.** *Cancer Res* 1994, 54:4813-4817.
17. Leach FS, Elledge SJ, Sherr CJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B: **Amplification of cyclin genes in colorectal carcinomas.** *Cancer Res* 1993, 53:1986-1989.

18. Arber N, Hibshoosh H, Moss SF, Sutter T, Zhang Y, Begg M, Wang S, Weinstein IB, Holt PR: **Increased expression of cyclin D1 is an early event in multistage colorectal carcinogenesis.** *Gastroenterology* 1996, **110**:669-674.
19. Holland TA, Elder J, McCloud JM, Hall C, Deakin M, Fryer AA, Elder JB, Hoban PR: **Subcellular localization of cyclin D1 protein in colorectal tumors is associated with p21 (WAF1/CIP1) expression and correlates with patient survival.** *Int J Cancer* 2001, **95(5)**:302-306.
20. Pasz-Walczak G, Kordek R, faflik M: **P21(WAF1) expression in colorectal cancer: correlation with p53 and cyclin D1 expression, clinicopathological parameters and prognosis.** *Pathol Res Pract* 2001, **197(10)**:683-689.
21. Utsunomiya T, Doki Y, Takemoto H, Shiozaki H, Yano M, Sekimoto M, Tamura S, Yasuda T, Fujiwara Y, Monden M: **Correlation of beta-catenin and cyclin D1 expression in colon cancers.** *Oncology* 2001, **61(3)**:226-233.
22. Maeda K, Chung YS, Kang SM, Ogawa M, Onoda N, Nakata B, Nishiguchi Y, Ikehara T, Okuno M, Sowa M: **Overexpression of cyclin D1 and p53 is associated with disease recurrence in colorectal adenocarcinoma.** *Int JCancer* 1997, **74**:310-315.
23. McKay JA, Douglas JJ, Ross VG, Curran S, Loane JF, Ahmed FY, Cassidy J, McLeod HL, Murray GI: **Analysis of key cell cycle checkpoint proteins in colorectal tumors.** *J Pathol* 2002, **196**:386-393.
24. Mate JL, Ariza A, Aracil C, Lopez D, Isamat M, Perez-Piteira J, Navas-Palacios JJ: **Cyclin D1 overexpression in non-small cell lung carcinoma: correlation with Ki-67 labeling index and poor cytoplasmic differentiation.** *J Pathol* 1996, **180**:395-399.
25. Buckley MF, Sweeney KJ, Hamilton JA, Sini RL, Manning DL, Nicholson RI, DeFazio A, Watts CK, Musgrove EA, Sutherland RL: **Expression and amplification of cyclin genes in human breast cancer.** *Oncogene* 1993, **8**:2127-2133.
26. Habermann J, Lenander C, Roblick UJ, Kruger S, Ludwig D, Alalya A, Freitag S, Dumbgen L, Bruch HP, Stange E, Salo S, Tryggvason K, Auer G, Schimmelpenning H: **Ulcerative colitis and colorectal carcinoma: DNA profile, laminin-5 gamma 2 chain and cyclin A expression as early markers for risk assessment.** *Scand J Gastroenterol* 2001, **36**:751-758.
27. Kubota Y, Petras RE, Easley KA, Bauer TW, Tubbe RR, Fazio VW: **Ki-67-determined growth fraction versus standard staging and grading parameters in colorectal carcinoma. A multivariate analysis.** *Cancer* 1992, **70**:2602-2609.
28. Shain AA, Ro JY, Brown RW, Ordonez NG, Cleary KR, El-Naggar AK, Wilson P, Ayala AG: **Assessment of Ki-67-derived tumor proliferative activity in colorectal adenocarcinomas.** *Mod Pathol* 1994, **7**:17-22.
29. Suzuki H, Matsumoto K, Terabe M: **Ki-67 antibody labeling index in colorectal carcinoma.** *J clin Gastroenterol* 1992, **15**:317-320.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-230X/4/22/prepub

British Journal of Cancer (2004) 90, 1612–1619

www.bjcancer.com

Correlation of amplification and overexpression of the c-*myc* oncogene in high-grade breast cancer: FISH, *in situ* hybridisation and immunohistochemical analyses

J Blancato*[1,2], B Singh[2], A Liu[3], DJ Liao[4] and RB Dickson[2,5]

[1]Institute for Molecular and Human Genetics, 3970 Reservoir Road, NW, Washington DC 20007, USA; [2]Lombardi Comprehensive Cancer Center, 3970 Reservoir Road, NW, Washington DC 20007, USA; [3]Biometry and Mathematical Statistics Branch, National Institute of Child Health and Human Development, 9000 Rockville Pike, Bethesda, MD 20892-7510, USA; [4]Department of Radiation Oncology, Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA; [5]Department of Oncology, Georgetown University Medical Center, 3970 Reservoir Road, NW, Washington DC 20007, USA

In this study, we analysed gene amplification, RNA expression and protein expression of the c-*myc* gene on archival tissue specimens of high-grade human breast cancer, using fluorescent *in situ* hybridisation (FISH), nonradioactive *in situ* hybridisation and immunohistochemistry. The specific question that we addressed was whether expression of c-Myc mRNA and protein were correlated with its gene copy amplification, as determined by FISH. Although c-*Myc* is one of the most commonly amplified oncogenes in human breast cancer, few studies have utilised *in situ* approaches to directly analyse the gene copy amplification, RNA transcription and protein expression on human breast tumour tissue sections. We now report that by using the sensitive FISH technique, a high proportion (70%) of high-grade breast carcinoma were amplified for the c-*myc* gene, irrespective of status of the oestrogen receptor. However, the level of amplification was low, ranging between one and four copies of gene gains, and the majority (84%) of the cases with this gene amplification gained only one to two copies. Approximately 92% of the cases were positive for c-*myc* RNA transcription, and essentially all demonstrated c-myc protein expression. In fact, a wide range of expression levels were detected. Statistically significant correlations were identified among the gene amplification indices, the RNA expression scores and protein expression scores. c-*myc* gene amplification, as detected by FISH, was significantly associated with expression of its mRNA, as measured by the intensity of *in situ* hybridisation in invasive cells ($P = 0.0067$), and by the percentage of invasive cells positive for mRNA expression ($P = 0.0006$). c-*myc* gene amplification was also correlated with the percentage of tumour cells which expressed high levels of its protein, as detected by immunohistochemistry in invasive cells ($P = 0.0016$). Thus, although multiple mechanisms are known to regulate normal and aberrent expression of c-*myc*, in this study, where *in situ* methodologies were used to evaluate high-grade human breast cancers, gene amplification of c-*myc* appears to play a key role in regulating expression of its mRNA and protein.
British Journal of Cancer (2004) **90**, 1612–1619. doi:10.1038/sj.bjc.6601703 www.bjcancer.com
Published online 30 March 2004

Keywords: c-*myc*; breast cancer; gene amplification; gene expression

Genetics and Genomics

The c-*myc* oncogene has been shown to be amplified and/or overexpressed in many types of human cancer (Marcu *et al*, 1992; Nass and Dickson, 1997; Nesbit *et al*, 1999; Liao and Dickson, 2000). Numerous experiments *in vivo* have also causally linked aberrant expression of this gene to the development and progression of cancer in different body sites (Marcu *et al*, 1992; Nass and Dickson, 1997; Nesbit *et al*, 1999; Liao and Dickson, 2000). However, several critical issues regarding the significance of c-*myc* in human cancer still remain obscure. First, even for a given type of malignancy, the frequencies of the alterations of c-*myc* at the cytogenetic and expression levels vary greatly from one report to another (Liao and Dickson, 2000). For instance, the frequencies of its amplification, mRNA and protein overexpression in breast cancer vary between 1–94, 22–95 and roughly 50–100%, respectively, among different reports (Liao and Dickson, 2000). Thus, it is still unclear to what extent this gene is altered at the cytogenetic level and at different expression levels in breast carcinoma.

One controversial issue pertains to the prognostic value of c-*myc* gene alterations in cancer. The central role of c-Myc protein in accelerating cell proliferation, documented by many early studies, has led to a general concept for many types of cancer that amplification or overexpression of this gene may be associated with a more aggressive tumour and a poorer patient survival (Berns *et al*, 1992; Marcu *et al*, 1992; Sato *et al*, 1995; Nass and Dickson, 1997; Nesbit *et al*, 1999; Visca *et al*, 1999; Liao and Dickson, 2000). However, many reports have shown an opposite correlation (Sikora *et al*, 1985, 1987; Watson *et al*, 1986; Polacarz *et al*, 1989; Voravud *et al*, 1989; Williams *et al*, 1990; Melhem *et al*, 1992; Pietilainen *et al*, 1995; Diebold *et al*, 1996; Smith and Goh, 1996; Augenlich *et al*, 1997; Bieche *et al*, 1999), while other studies do not support either of these conclusions. For instance, gene amplification or overexpression of c-Myc protein has also been shown to associate with a better tumour differentiation or a better

*Correspondence: Dr J Blancato; E-mail: blancatj@georgetown.edu
Received 7 July 2003; revised 17 January 2004; accepted 17 January 2004; published online 30 March 2004

patient survival for cancer of the testis, ovary, bile ducts, colon and breast (Sikora *et al*, 1985, 1987; Watson *et al*, 1986; Polacarz *et al*, 1989; Voravud *et al*, 1989; Williams *et al*, 1990; Melhem *et al*, 1992; Pietilainen *et al*, 1995; Diebold *et al*, 1996; Smith and Goh, 1996; Augenlich *et al*, 1997; Bieche *et al*, 1999). This controversy does not appear to be related completely to the cancer type, since both positive (Berns *et al*, 1992; Visca *et al*, 1999) and negative (Williams *et al*, 1990; Melhem *et al*, 1992; Pietilainen *et al*, 1995; Smith and Goh, 1996; Augenlich *et al*, 1997; Bieche *et al*, 1999) correlations have been reported for colon cancer and breast cancer. More interestingly, c-Myc overexpression has been shown to predict a poorer prognosis for cutaneous melanoma, but a favourable outcome for uveal melanoma (Grover *et al*, 1997; Chana *et al*, 1998a, b, 1999; Grover *et al*, 1999). These data indicate different roles of c-Myc, even in the same type of tumour, perhaps depending upon different tissue microenvironments.

Another controversial issue concerns the nuclear–cytoplasmic localisation of c-Myc. Studies of neoplasms of the colon, testis, ovary and liver have shown that predominantly nuclear localisation of c-Myc tends to occur in benign lesions, while cytoplasmic localisation tends to occur in more malignant tumours (Sikora *et al*, 1985; Sundaresan *et al*, 1987; Melhem *et al*, 1992; Sasano *et al*, 1992; Yuen *et al*, 2001). Whether these patterns of subcellular localisation of c-Myc tend to reflect the malignant status of breast cancer remains an enigma.

A recent study of the impact of DNA amplification on gene expression patterns in breast cancer used mRNA and DNA from 14 breast cancer cell lines. Analysis was conducted with a 13 000 cDNA clone array for gene expression measurement and a Comparative Genomic Hybridisation (CGH) microarray for gene copy number measurements. This study also included known breast cancer genes, such as c-*myc*, *HER/2-neu* and *aib1* (Hyman *et al*, 2002). Interestingly, 44% of the most highly amplified genes were also overexpressed at the mRNA level. Consistent with this pattern, c-*Myc* gene copy number and its expression levels showed a statistically significant ($\alpha = 0.020$) correlation in this microarray study of breast cancer cell lines. Another study, by Pollack and colleagues, used microarray analysis and BAC array CGH of RNA and DNA (respectively) extracted from intermediate grade human breast tissues, and tested for amplification and expression of c-Myc (among other genes). This study demonstrated that two out of 37 specimens were both amplified and overexpressed, while others were either amplified or overexpressed, but not both. The authors of this study suggested that contaminating stromal tissue may compress the fluorescence ratios leading to underestimates of gene amplification and overexpression (Pollack *et al*, 2002).

To more clearly address the importance of gene amplification and expression of c-Myc in human breast cancer, we used *in situ* methodologies, which can clearly distinguish stromal and carcinoma components. We studied the amplification and over expression of the c-*myc* gene with fluorescent *in situ* hybridisation (FISH), non-radioactive *in situ* hybridisation (ISH) and immunohistochemical (IHC) approaches on paraffin-embedded biopsy sections of untreated, high-grade breast cancer. It was observed that 70, 92 and 70% of the cancer cases exhibited c-*myc* gene amplification, its mRNA overexpression and its protein over expression, respectively. In most of the cases (84%) that showed gene amplification, the c-*myc* gene gained only one to two copies, which is consistent with c-*myc* FISH data from other studies. Unlike some oncogenes, such as N-*myc*, which typically demonstrates gene amplification copy numbers of greater then 10 in neuroblastoma, and *HER-2/neu* (Sartelet *et al*, 2002), whose copy numbers range up to 14–40 in breast carcinomas (Isola *et al*, 1999), gene copy numbers of c-*myc* are not as greatly increased. In the study noted earlier, using breast cancer cell line CGH array and cDNA microarray expression analysis, it was demonstrated that the most dramatically increased expression levels were associated with large gene copy number increases, although low-level gains and losses had a significant influence on gene expression dysregulation (Hyman *et al*, 2002). Only one study has been published (Pollack *et al*, 2002) that has begun to determine if these findings are directly relevant to actual human breast tumour tissues, since many of the genetic changes in tissue culture cell lines are more extreme than those displayed in primary tumour material. Furthermore, the relationships among gene amplification, mRNA expression and c-Myc protein expression were not explored in prior human breast cancer cell line and tumour tissue studies (Hyman *et al*, 2002; Pollack *et al*, 2002).

In our human breast tumour tissue study, a high correlation was found between c-*myc* FISH and ISH, for both percentage of staining ($P<0.0067$) and intensity positive cells ($P<0.0006$). In addition, c-*myc* gene copy amplification by FISH was correlated with c-Myc protein expression positive cells by IHC ($P<0.0016$). These results support the idea that c-Myc overexpression of both mRNA and protein is related to the copy number of the *c-myc* DNA amplification. We show in this study that amplification and overexpression of c-Myc occur with high frequency in high-grade human breast cancer tissues.

MATERIALS AND METHODS

Materials

Formalin-fixed, paraffin-embedded tissue blocks of breast carcinoma and normal breast tissue were obtained from the Histopathology and Tissue Shared Resource at the Lombardi Comprehensive Cancer Center (LCCC), at Georgetown University Medical Center. The criteria for tumour selection were the following: negative progesterone receptor status, metastases to auxiliary lymph nodes and high grade (Elston Score >7). The oestrogen receptor status of the tumours was known from archived pathology reports. The parameters were chosen from our prior meta-analysis (Deming *et al*, 2000), as indications of a high likelihood of c-*myc* gene amplification. Normal breast tissue specimens were from reduction mammoplasty. Serial sections (5 μm) for FISH, ISH and IHC were prepared by the LCCC Histopathology and Tissue Shared Resource.

FISH

A dual-label FISH technique was used (Jenkins *et al*, 1997). Slides were baked overnight at 60°C to assure adherence of the sample. Tissue sections were deparaffinised with two successive, 10 min xylene washes, and then dehydrated in a graded ethanol series of 70, 80 and 95% at room temperature. Samples were then digested with 4% pepsin (Sigma, St Louis, MO, USA) at 45°C for 10 min. DNA probes used were an alpha satellite probe to chromosome 8, labelled with biotin, and a *c-myc* probe, labelled with digoxigenin (Ventana, Tucson, AZ, USA). Codenaturation was performed at 90° for 10 min on a hot plate. Hybridisation was at 37°C for 12–16 h. Detection of signals was accomplished with an antiavidin antibody labelled with Texas Red, and an antidigoxigenin antibody conjugated to fluorescein (Ventana, Tucson, AZ, USA). Slides were postwashed in 2 × SSC at 72°C for 5 min and counterstained with DAPI to visualise cell nuclei. Results were viewed and quantified with a Zeiss Axiophot fluorescence microscope, equipped with appropriate filters and an Applied Imaging Cytovision system (Pittsburgh, PA, USA). In this approach, the c-*myc* unique sequence probe was visualised as a green signal and the control probe for the chromosome 8 centromere was red, thus easily being distinguished when scored.

One serial section from each tumour sample was stained with haematoxylin and eosin and first reviewed by a pathologist (BS), to help identify the tumour area of the section. This procedure ensured that the tumour cells, but not the normal cells, were

counted. Nuclei of up to 50 tumour cells were scored from each FISH-stained section, independently by two investigators. Hybridisation signals were averaged, and the amplification index was presented as the number of c-*myc* signals divided by the number of chromosome 8 centromere signals. A 1.8-fold increase was used as the criterion to judge the presence of c-*myc* gene amplification.

In situ hybridisation

In situ hybridisation (ISH) was carried out with a nonradioactive method, described previously (Liao *et al*, 2000a, b). One serial section from each specimen was hybridised overnight at 60°C with riboprobes, that were *in vitro* transcribed from the antisense or sense strand of an approximately 300 bp cDNA of human c-*myc* (ATCC, Manassas, VA, USA), labelled with digoxigenin-conjugated UTP. The sections were then incubated with an antibody against digoxigenin, followed by incubation with a second antibody conjugated to alkaline phosphatase. The signal was visualised by colour development with 5 bromo-4-chloro-3-indolyl phosphate and nitroblue tetrazolium. All reagents were purchased from Boehringer Mannheim, Indianapolis, IA. To control the signal specificity, two serial sections were mounted on the same slide for hybridisation with the antisense and sense probes, respectively. ISH was given an intensity and percentage scores, based on intensity of positive staining and number of cells staining, respectively. Intensity scores were assigned 0, 1, 2 and 3, and percentage scores were assigned as 1- 1–25, 2- 26–50, 3- 51–75 and 4- 76–100%.

Immunohistochemistry

Immunohistochemical staining (IHC) was performed using an avidin–biotin complex (ABC) method described previously (Liao *et al*, 1998). One serial section of each specimen was deparaffinised and blocked with 3% peroxide. Antigens were retrieved by heating slides in a microwave oven in 50 mM citrate buffer, pH 6.4, at boiling temperature, for 12 min. After blocking with 6% normal goat serum, the section was incubated with a mouse monoclonal antibody to human c-Myc (9E10, Sigma Chemical Company, St Louis, MO, USA) at 1:100 dilution for 2 h, followed by 1 h incubation with a second antibody conjugated with biotin (Vector Laboratories Inc., Burlingame, CA, USA). The section was then incubated with peroxidase-conjugated avidin (Dako, Corporation, Carpinteria, CA, USA) for 30 min, followed by colour development with diaminobenzidine and peroxide. All procedures were carried out at room temperature. To control the signal specificity, serial sections from 10 tumour samples were also stained using an alternate c-Myc antibody (C19 from Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) at 1:60 dilution. This antibody resulted in focally positive staining in the tumour, but the staining intensity was weaker. To control the signal specificity, serial sections were made from five selected positive cases which were subjected to the same staining procedure, with a normal mouse IgG to replace the c-Myc antibody. This control staining did not give rise to a signal, demonstrating the specificity of the c-Myc antibody signal. IHC staining was given an intensity and percentage score based upon the intensity of positive staining and number of cells staining. Intensity scores were assigned 0, 1, 2 and 3 and percentage scores were assigned as 1- 1–25, 2- 26–50, 3- 51–75 and 4- 76–100%. Determinations were made of cellular localisation of c-Myc antibody staining to cytoplasm and/or nucleus in normal and invasive cells within each breast tumour specimen.

Statistical analyses

For each analysis of gene copy amplification (FISH), mRNA expression (ISH) and protein expression (IHC), all cases were first grouped as positive or negative to calculate the percentages of positive cases and negative cases, as described (Zar, 1974). Fisher's exact test was used to compare percentages, and two-sample *t*-test or Wilcoxon rank test was used to compare average scores. Both ISH and IHC were given intensity and percentage scores, based on intensity of positive staining and number of cells staining, respectively. As noted earlier, intensity scores were assigned 0, 1, 2 and 3 and percentage scores were assigned as 1- 1–25, 2- 26–50, 3- 51–75 and 4- 76–100%. A score of >2 for either intensity of staining or percentage of cells positive by ISH was assigned as high. For IHC, an intensity score of >1 was assigned as high and a percentage score of >3 was categorised as high. Each amplification index was paired with its corresponding mRNA expression score to calculate the coefficient *r*. The same method was used to estimate the association of the amplification indices with the c-Myc protein expression levels, and the association of the mRNA expression levels with the protein expression levels. A *P*-value of 0.05 or less was used to determine the statistical significance in all analyses. In all, 54 pairs of normal *vs* invasive tissues were analysed using McNemars χ^2 test to determine if there was a difference in cellular localisation of c-Myc antibody signal to nuclear or cytoplasmic compartments.

RESULTS

FISH analysis of gene amplification

Amplification of the c-*myc* gene was measured by a FISH test in 46 cases of breast cancer; Figure 1 demonstrates cells with no amplification (one copy of c-*myc* /one copy of chromosome 8 centromere, and a moderate amplification a 3/1 ratio). Amplification was calculated by the number of c-*myc* signals divided by the number of chromosome 8 alpha satellite signals. A 1.8-fold increase cut-off was used to judge gene amplification. As shown in Table 1, 32 out of 46 (70%) cases were gene amplified for c-*myc*, whereas only 30% (14/46) of the cases showed amplification indices lower than the cut-off value. The amplification indices for most (84%, or 27/32) cases with gene amplification, ranged between 1.8- and three-fold, indicating that the locus gained up to two copies of c-*myc* in the majority of the cases. The percentage of cases with gene gains of three copies or higher was 11% (five out of 46) of total cases analysed, or near 16% (five out of 32) of the cases with gene amplification, including one case (2% of total cases or 3% of the cases with gene amplification) with the highest index of 5 (a gain of four copies).

In all, 28 of the breast carcinomas in this study were ER negative, and 14 were ER positive. The average c-*myc* gene amplification score was 1.896 (s.e. = 0.196) for ER positive and 2.201 (s.e. = 0.157) for ER negative. Although ER-negative tumours had a slightly higher average c-*myc* score, the difference was not statistically significant (two-sided $P = 0.252$ from two-sample *t*-test and 0.251 from Wilcoxon rank test), consistent with the results of our prior meta-analysis of the literature (Deming *et al*, 2000).

In situ hybridisation analysis of c-*myc* mRNA expression

A total of 51 breast cancer samples were studied for c-Myc mRNA expression, with non radioactive *in situ* hybridisation (ISH). ISH results were assigned intensity and percentage scores based upon signal intensity of positive staining and number of cells staining within the sample, respectively. As shown in Table 2, 86% (44 out of 51) tumours were scored as high in intensity, and 92% (47out of 51) had more than 51% positive cells, also considered as highly increased c-Myc expression. mRNA expression was heterogeneous in the breast tumour tissue, and no morphologic subtype was predominant in the high or low categories. One case showed no c-Myc ISH staining. In 79% (38/48) of cases, epithelia in normal mammary glands adjacent to the tumour also showed a high

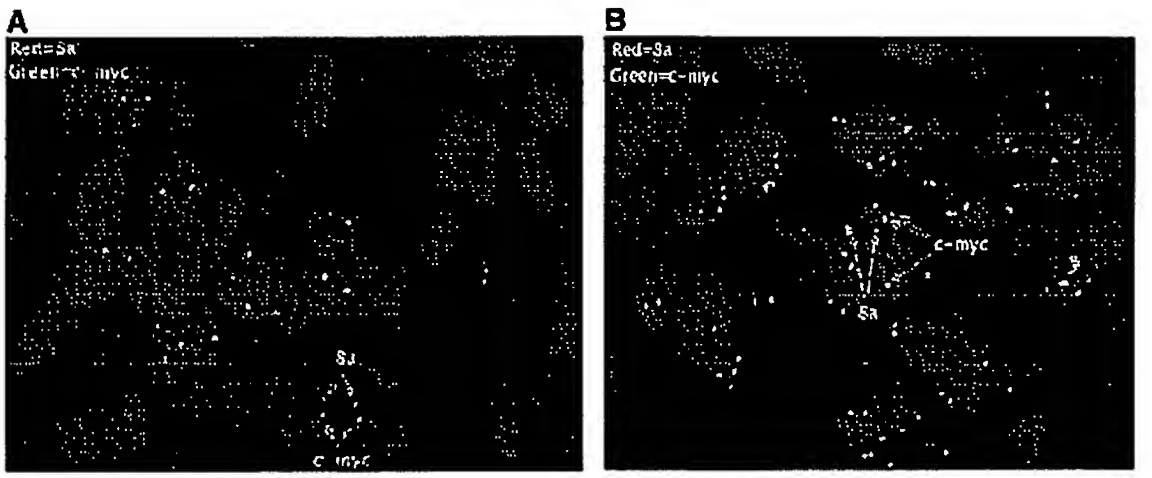


Figure 1 FISH analysis of c-*myc* amplification in tumour cells from breast tumour tissue sections. FISH probe for human c-*myc* unique-sequence is seen as green, while the normal control signal, a centromeric probe signal for chromosome 8 is shown in red. The nuclei of tumour cells were visualised by DAPI counter-staining. (**A**) 1 : 1 copy ratio of c-*myc* to chromosome 8 (c-*myc*/8 centromere), indicating no amplification of c-*myc* in tumour cells. (**B**) 1 : 3 copy ratio of c-*myc* to chromosome 8 (c-*myc*/8 centromere), a moderate amplification of the c-*myc* gene.

Table 1 c-*myc* gene copy amplification analysis by FISH in poor prognosis human breast tumour samples

Amplification index (#c-*myc* signals/# control signals)+	Percentage of samples with FISH ratios in each category	
1.0–1.7	30%	14 out of 46
1.8–1.99	20%	Nine out of 46
2.0–2.9	39%	18 out of 46
>3.0	11%	Five out of 46

Analysis was conducted on 46 individual paraffin-embedded tissue samples with negative progesterone receptor status, positive lymph node involvement and high tumour grade. +Normal control ratio is 1.

intensity of staining. In three cases, no staining was seen in the normal terminal duct lobular units. Figure 2 shows representative fields of high, medium and low *c-myc* mRNA expression levels in invasive ductal carcinoma samples.

Association of FISH and ISH

c-Myc scores were dichotomised as binary variables (high or low), and a score of 2 or higher was categorised as high on ISH. A score higher than median was categorised as high from FISH studies. These dichotomised scores are depicted in Table 3. A Fisher's exact test was performed for comparing binary responses to see if there was any association between FISH and ISH. It was found that the FISH score was significantly associated with percentage of staining in the invasive cells ($P=0.0067$, two-sided McNemar's test) and also with the intensity score on ISH ($P=0.0006$, two-sided).

Immunohistochemical staining of c-Myc proteins

In total, 51 breast carcinomas, which were subjected to FISH analysis, and all of which also had been analysed for c-*myc* mRNA by *in situ* hybridisation, were also analysed for the expression of c-Myc protein, using immunohistochemical staining with the 9E10 antibody. IHC results were assigned an intensity and percentage score based on intensity of positive staining and number of cells staining, respectively. Intensity scores were assigned 0, 1, 2 and 3 and percentage scores were assigned as 0, 1- 0–25, 2- 26–50, 3- 51–75 and 4- 76–100. For IHC, an intensity score of >1 was assigned as high and a percentage score of >3 was categorised as high. Figure 2 shows examples of high, medium and low levels of c-myc antibody staining in invasive ductal carcinoma samples. In 34 cases, normal tissue was seen; 30 of these showed cytoplasmic staining and 22 had nuclear staining in terminal ductal lobular units. In all, 12 cases showed 1+, 14 cases 2+ and four cases 3+ cytoplasmic staining. *In situ* hybridisation revealed positive staining in 46 out of 49 cases with normal tissue. Seven cases showed 1+, 13 cases showed 2+ and 26 cases showed 3+ staining by ISH. Both immunohistochemistry and *in situ* hybridisation showed diffuse positivity in adipocytes.

Table 4 shows the staining pattern for the cohort. In all, 70% (36 out of 51) of cases showed high intensity of staining for c-Myc protein, while 85% (29 out of 34) of cases with detectable staining had more than 76% positive cells, also considered as high expression. To verify the staining specificity, serial sections from 10 tumour specimens that were positive for 9E10 antibody were also stained using the C19 rabbit polyclonal anti-c-Myc antibody. Results revealed a staining pattern similar to 9E10. However, the staining intensity with C19 was weaker than 9E10. The specificity of these two antibodies was verified by Western blots in previous studies (Persons *et al*, 1997; Liao *et al*, 2000b). Figure 2 shows results of c-Myc *in situ* hybridisation and immunohistochemistry studies on samples considered to demonstrate low, moderate and high levels of c-Myc expression. Analysis of c-Myc protein localisation results in the nucleus or cytoplasmic compartments of normal and invasive cells within the tumours revealed that nuclear staining was positive in 41% of normal cells, compared to 22% of invasive cells (statistical significance at $P=0.01$ by McNemar's two-sided χ^2 test). The increase in relative cytoplasmic localisation of c-Myc protein, comparing normal (53.7%), to invasive cells (61.1%) was not significantly different. Thus, the data are consistent with partial exclusion of c-Myc from the nuclei of invasive breast cancer cells.

The FISH score was significantly associated with the percentage positivity of invasive cells, as seen on IHC studies of c-Myc. However, 40% of tumours displayed a low index of c-*myc* gene amplification, but still expressed high levels of c-Myc protein (Table 6), indicating the possibility of other mechanisms of over expression unrelated to gene amplification in at least some tumours. The FISH score was not significantly associated with the intensity of IHC staining in the invasive cells (not shown), in contrast to the IHC percentage positivity score.

Table 2 c-myc mRNA *in situ* hybridisation (ISH) results

Staining Intensity	0	1	2	3	Percent positivity	1	2	3	4
Number of tumour samples in each category $N=51$	1	6	25	19	Number of tumour samples in each level category $N=51$	1	3	5	42

In all, 51 human high-grade breast carcinomas were analysed to determine the relationships between c-Myc mRNA expression and *c-myc* gene *in situ* hybridisation results. Data are shown in two ways in the above table. First, overall staining intensity of c-Myc-positive cells was scored as 0, 1, 2, 3 (low to high), and the number of tumour samples at each level of staining indicated on the line below. Next, the percentage of tumour cells staining was scored as 0, 1, 2, 3, 4 (low to high %, as discussed in Materials and Methods). The number of tumours at each level of percent cell positivity for c-Myc is then indicated on the line below.

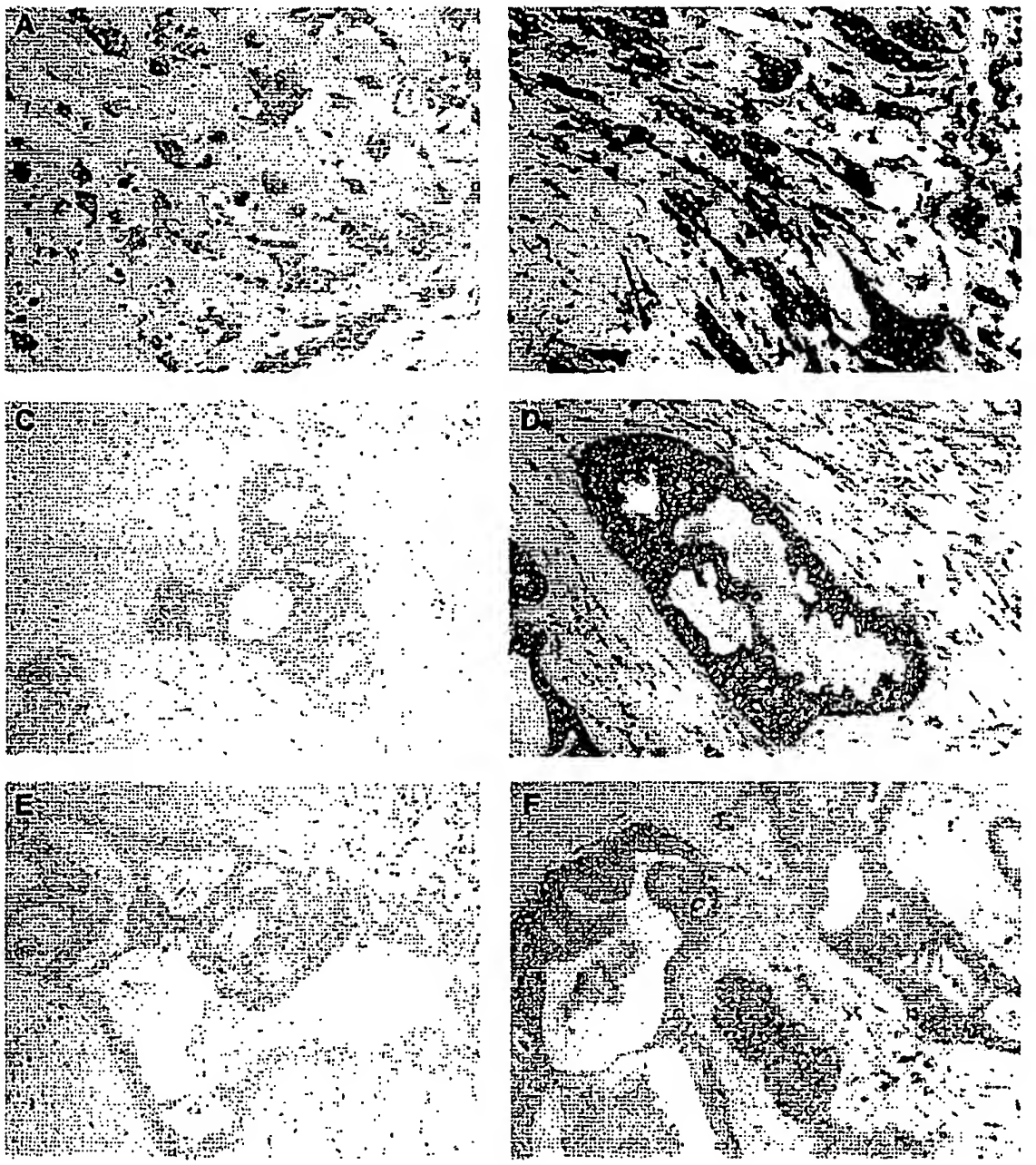


Figure 2 Immunohistochemical staining and *in situ* hybridisation for c-Myc of three sets of invasive ductal carcinoma. (**A**, **C** and **E**) High (3+), intermediate (2+) and low (1+) level of staining by immunohistochemistry for c-Myc. (**B**, **D** and **F**) High (3+), intermediate (2+) and low (1+) level of staining by *in situ* hybridisation.

DISCUSSION

Although there have been many reports on c-*myc* amplification in human breast cancer (Liao and Dickson, 2000), there are only two published studies involving application of the FISH technique to unfixed, frozen sections (Persons *et al*, 1997; Visscher *et al*, 1997), and one prior study using FISH on an archival human tissue microarray (Schraml *et al*, 1999). Another recent study applied FISH to evaluate c-*myc* amplification in ductal carcinoma *in situ* (DCIS) (Aulmann *et al*, 2002). Using the FISH technique on formalin-fixed, paraffin-embedded sections, we now show that 70% of high-grade breast cancer samples bear c-*myc* gene copy amplifications. Interestingly, the above-mentioned study, using FISH and focusing on DCIS, detected amplification of c-*myc* in only 20% of cases, but found a correlation of c-*myc* with increased tumour size and proliferation (Aulmann *et al*, 2002).

The level of amplification of c-*myc* in our study ranged between one and four additional copies of the gene; the majority (84%) of

the cases with the gene amplification gained only one to two copies, also consistent with FISH data reported for c-*myc* copy amplification in human metastatic prostate carcinoma tissues (Jenkins *et al*, 1997). The relationship between the level of c-*myc* gene copy amplification and the level its increased mRNA expression has been examined previously in breast cancer cell lines (Hyman *et al*, 2002). In general, it has been concluded that the two scores coordinate for c-*myc*, as is the case for many breast cancer genes. However, only 44% of the highly amplified genes, in general, showed increased RNA expression, and only 10.5% of the highly overexpressed genes were gene copy-amplified in the cell line study (Hyman *et al*, 2002). Another analysis was conducted to study of relationships between gene amplification and expression of 6095 genes in 37 intermediate grade human breast tumours. This study demonstrated that 62% of the highly amplified genes also showed elevated expression; overall, a two-fold change in DNA copy number was associated with a 1.5-fold change in mRNA levels. Overall, 12% of the variation in gene expression in the breast tumours studied was associated with gene copy number variation (Pollack *et al*, 2002). Further study of additional human breast tumours, at precisely defined grades and stages, will be necessary in order to more fully define the relationships between DNA copy numbers and expression of genes. The studies we report here indicate higher levels of c-Myc gene amplification and expression, than other previous reports in breast cancer. We believe that this is probably the result of our analysis of individual tumour cells in a well-defined set of high-grade breast tumours. Prior c-Myc expression and amplification microarray studies used tumour specimens which contain normal stromal components, potentially underestimating amplification and expression levels of the invasive tumour components (Pollack *et al*, 2002).

Our study reports a percentage of tumours gene amplified for c-*myc* (using FISH in high-grade tumours) that is much higher than the average figure (15.5%) reported in the literature (Isola *et al*, 2002). Most of the prior studies have employed the relatively insensitive Southern blot technique, and were reviewed in a recent meta-analysis (Deming *et al*, 2000). Consistent with this prior literature background, a recent study of 94 lobular and ductal breast cancers assessed amplification of c-*myc* by using a semiquantitative PCR assay and protein expression, with

Table 5 Nuclear/cytoplasmic localisation of c-Myc comparing normal and invasive cells

Normal cells (frequency percent)	Invasive cells (frequency percent)		Total
(A) Nuclear localisation			
	−	+	
−	28	4	32
+	14	8	22
Total	42	12	54
(B) Cytoplasmic localisation			
−	12	13	25
+	9	20	21
Total	21	33	54

In all, 54 pairs (normal vs invasive) of tissues were analysed to answer the questions of (1) whether positivity of nuclear cells in normal tissues is different from that in invasive cells, and similarly (2) whether positivity of cytoplasmic cells in normal tissues is different from that in invasive cells. The data are summarised in the above contingency tables. In all, 22 normal cell specimens were positive for c-Myc staining (40.71%), compared to 12 specimens (22.2%) in invasive cells. The difference is statistically significant ($P = 0.01$) by McNemar's χ^2 test (two-sided).

Table 3 Correlations between c-*myc* gene copy number (FISH) mRNA expression (ISH)

	FISH	
	Low	High
(A) ISH (% cells)		
Low	*1*	3
High	*19*	*18*
		$P = 0.0067$
(B) ISH (intensity)		
Low	*2*	*5*
High	*18*	*16*
		$P = 0.0006$

Serial sections of high-grade human breast carcinomas were scored for c-*myc* gene copy number (FISH, Table 1) and mRNA expression (ISH, Table 2). In (A), a positive correlation ($P = 0.0067$) was observed between tumour samples with a high percentage of cells demonstrating mRNA expression and a high c-*myc* gene copy number. A score of 2 or higher was classified as high on ISH, and a score of median or greater was categorised as high on FISH. In (B), a positive correlation ($P = 0.0006$) was shown between a high level of intensity for c-Myc RNA expression a high and c-*myc* gene copy number. Note that a pairwise comparison of FISH and ISH was not possible for all cases, due to incomplete overlap of cases analysed with each assay.

Table 6 Correlation between c-Myc protein expression (IHC) and c-*myc* gene copy number (FISH)

	FISH	
IHC (% cells)	Low	High
Low	3	0
High	10	15
		$P = 0.0016$

Consecutive serial sections of high-grade human breast tumours were scored for c-*myc* gene copy number or protein expression, by immunohistochemistry (IHC). IHC scores were defined in the Materials and methods section. Data were analysed for correlations between the results. A highly significant correlation was observed between high c-Myc protein expression (IHC) between percent cells positive and high c-*myc* gene amplification (FISH). $P = 0.0016$ from two-sided McNemar's test. Note that for 15 cases, no staining for c-Myc could be detected; these negative cases were not included in the correlation presented, above.

Table 4 c-Myc immunohistochemistry (IHC) results

Staining intensity	0	1	2	3	Percent positivity	1	2	3	4
Number of tumour samples in each category	15	13	20	3	Number of tumour samples in each category	2	2	1	29

In all, 51 high-grade human breast carcinomas were analysed to determine the relationships between c-Myc protein expression and c-*myc* gene *in situ* hybridisation results. Data are shown in two ways in the above table. First, overall staining intensity of c-Myc-positive cells was scored as 0, 1, 2, 3 (low to high), and the number of tumour samples at each level of staining is indicated on the line below. Next, in a random subset of these cases, the percentage of tumour cells staining was scored as 0, 1, 2, 3, 4 (low to high %, as discussed in Materials and methods). The number of tumours at each level of percent cell positivity for c-Myc is indicated on the line below.

densitometry, after Western blot. These data showed c-*myc* gene amplification in 21% of tumours (Jenkins *et al*, 1997), using assays not based on *in situ* discrimination of tumour *vs* nontumour cells. The lower frequency of c-*myc* in this prior study is in contrast with the data we present here, and could be the result of the higher sensitivity and precision of the FISH and immunohistochemical methods, as distinct from quantitative PCR and Western blot densitometry. In addition, the 70% of amplified tumours in our study is also much higher than the 12% reported by Schraml *et al* (1999), using a c-*myc* FISH test on a tissue microarray. This large difference may be because the arrays are prepared from cores of paraffin-embedded tissue, as small as 0.6 mm in diameter which may contain too few tumour cells for complete analysis of amplification of a gene, such as c-*myc*. c-*myc* is known to be quite heterogeneous in its gene amplification within individual tumours (in contrast to *HER2/neu*, for example) (Persons *et al*, 1997).

Most previous reports on the expression of c-*myc* mRNA have utilised Northern blot, dot blot or PCR-based approaches, while just a few involved *in situ* hybridisation, which were primarily performed on frozen tissue sections (Liao and Dickson, 2000). Normal breast tissue is dominated by adipose cells, differing greatly from tumour tissue in its epithelial cellularity. Thus, normal and tumour tissues may not be rigorously compared by techniques involving RNA extraction from total tissue. Therefore, conclusions such as 'increased expression' may be more difficult to make from studies with Northern blot, dot blot and PCR-based techniques that require RNA extraction from tissues that have not been fastidiously micro-dissected for selection of tumour cells. Using a more sensitive, nonradioactive *in situ* hybridisation (ISH) approach on formalin-fixed, paraffin-embedded sections, we report herein high expression of c-*myc* mRNA in 92% of high-grade breast carcinomas. This figure is much higher than the recently reported data (22%), obtained by using a real-time RT–PCR method (Bieche *et al*, 1999). Dilution of the RNA from epithelium by the RNA from adipose in normal breast tissue in this latest prior report may be one of the possible explanations for this large difference.

In conclusion, the present study shows that approximately 70, 92 and 70% of biopsies of untreated high-grade breast cancer exhibit c-*myc* gene amplification, mRNA overexpression and protein overexpression, respectively. In most cases (84%), with gene copy amplification, the c-*myc* gene gains one to two additional copies. c-*myc* gene amplification was significantly associated with expression of its mRNA (both by intensity in invasive cells and by percentage positivity in invasive cells), and with expression of its protein (by percentage positivity in invasive cells). However, our data were also consistent with the prior literature on c-Myc (reviewed in Nass and Dickson, 1997; Liao and Dickson, 2000), indicating complex transcriptional, post transcriptional, translational and post-translational control of c-Myc expression *in vitro*. Specifically, in Table 5 we observed that in 40% of the high-grade tumours tested, c-Myc protein was expressed at high levels, despite a lack of its gene amplification.

It will be interesting to analyse lower grade tumours and premalignant lesions, with the same measurement tools, to determine if this c-*myc* amplification pattern is different, comparing different steps in onset and progression of the disease. Specifically, prior studies in fibroblasts and in human mammary epithelial cells (Liao *et al*, 1998, 2000a, b) have demonstrated that only a subtle deregulation of expression of c-Myc is sufficient to allow genomic instability. These prior cell biologic findings raise the question of whether c-Myc protein expression precedes or follows its gene amplification during the course of the natural history of breast cancer. It will also be interesting for future studies of lower grade breast cancers and premalignant lesions to determine whether there is evidence of nuclear exclusion of c-Myc protein. Indeed, nuclear exclusion of c-Myc in high-grade tumours could serve to attenuate its functions in later stages of disease progression (Liao and Dickson, 2000).

ACKNOWLEDGEMENTS

We would like to express our gratitude to Mr G Veytsman and H-K Cao for scoring FISH data, Mrs S Constable for helping in tissue sectioning. This work was supported by NIH Grants RO1 CA72460 and AG1496 to RB Dickson and by a pilot grant from the Georgetown University ACS Institutional Research Grant to JK Blancato. This work was supported by NIH Grants RO1 CA72460 and AG1496 to RB Dickson and by an ACS pilot grant (from the GU Institutional ACS Research Grant) to JK Blancato.

REFERENCES

Augenlich L, Wadler S, Gorner G, Richards C, Ryan L, Multani A, Pathak S, Benson A, Haller D, Heerdt B (1997) Low-level c-*myc* amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Res* 57: 1769–1775

Aulmann S, Bentz M, Sinn HP (2002) C-myc oncogene amplification in ductal carcinoma *in situ* of the breast. *Breast Cancer Res Treat* 74: 25–31

Berns E, Klijn J, Van Putten W, van Staveren I, Portengen H, Foekens J (1992) c-*myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res* 52: 1107–1113

Bieche I, Laurendeau I, Tozlu S, Olivi M, Vidaud D, Lidereau R, Vidaud M (1999) Quantitation of MYC gene expression in sporadic breast tumors with a real-time transcription-PCR assay. *Cancer Res* 59: 2759–2765

Chana J, Cree I, Foss A, Hungerford J, Wilson G (1998a) The prognositc significance of c-*myc* oncogene expression in uveal melanoma. *Melanoma Res* 8: 139–144

Chana J, Grover R, Wilson G, Hudson D, Forders M, Sanders R, Grobbelaar A (1998b) The clinical significance of c-*myc* oncogene expression in melanomas of the scalp. *Br J Plast Surg* 51: 191–194

Chana J, Wilson G, Cree I, Alexander R, Myatt N, Neale M, Foss A, Hungerford J (1999) c-*myc*, p53, and *bcl-2* expression and clinical outcome in uveal melanoma. *Br J Opthamol* 83: 110–114

Deming S, Nass S, Dickson R, Trock B (2000) c-*myc* amplification in breast cancer: a meta-analysis of its occurrence and prognostic relevance. *Br J Cancer* 83: 1688–1695

Diebold J, Suchy B, Baretton G, Blasenbreu S, Meier W, Schmidt M, Rabes H, Lohrs U (1996) DNA ploidy and MYC DNA amplification in ovarian carcinomas: correlation with p53 and bcl-2 expression, proliferative activity and prognosis. *Virchows Arch* 429: 221–227

Grover R, Chana J, Grobbelaar A, Hudson D, Forders M, Wilson G, Sanders R (1999) Measurement of c-*myc* oncogene expression provides an accurate prognostic marker for acral lentiginous melanoma. *Br J Plast Surg* 52: 122–126

Grover R, Ross D, Wilson G, Sanders R (1997) Measurement of c-*myc* oncoprotein provides an independent prognostic marker for regional metastatic melanoma. *Br J Plast Surg* 50: 478–482

Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi O, Kallioniemi A (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 62: 6240–6245

Isola J, Chu L, De Vries S, Matsumura K, Chew K, Ljung B, Waldman F (1999) Genetic alterations in ERBB2-amplified breast carcinomas. *Clin Cancer Res* 5: 4140–4145

Jenkins R, Qian J, Lieber M, Bostwick D (1997) Detection of c-*myc* oncogene amplification and chromosomal anomalies in metastatic

prostatic carcinoma by fluorescence *in situ* hybridization. *Cancer Res* 57: 524–531

Liao D, Dickson R (2000) c-Myc in breast cancer. *Endocrine-Relat Cancer* 7: 143–164

Liao D, Hou X, Bai S, Li S, Li J (2000a) Unusual deregulation of cell cycle components in early and frank estrogen-induced renal neoplasias in the Syrian hamster. *Carcinogenesis* 21: 2167–2173

Liao D, Natarajan G, Deming S, Jamerson M, Johnson M, Chepko G, Dickson R (2000b) Cell cycle basis for the onset and progression of c-Myc-induced, TGFa-enhanced mouse mammary gland carcinogenesis. *Oncogene* 19: 1307–1317

Liao D, Pantazis C, Hou X, Li S (1998) Promotion of estrogen-induced mammary gland carcinogenesis by androgen in the male Noble rat: probable mediation by steroid receptors. *Carcinogenesis* 19: 2173–2180

Marcu K, Bossone S, Patel A (1992) Myc function and regulation. *Annu Rev Biochem* 61: 809–860

Melhem M, Meisler A, Finley G, Bryce W, Jones M, Tribby I, Pipas J, Koski R (1992) Distribution of cells expressing Myc proteins in human colorectal epithelium, polyps, and malignant tumors. *Cancer Res* 52: 5853–5864

Nass S, Dickson R (1997) Defining a role for c-Myc in breast tumorigenesis. *Breast Cancer Res Treat* 44: 1–22

Nesbit C, Tersak J, Prochownik E (1999) MYC oncogenes and human neoplastic disease. *Oncogene* 18: 3004–3016

Persons D, Borelli K, Hsu P (1997) Quantitation of *HER-2/neu* and *c-myc* gene amplification in breast carcinoma using fluorescence *in situ* hybridization. *Mod Pathol* 10: 720–727

Pietilainen T, Lipponen P, Aaltomaa S, Eskelinen M, Kosma V, Syrjanen K (1995) Expression of *c-myc* proteins in breast cancer as related to established prognostic factors and survival. *Anticancer Res* 15: 959–964

Polacarz S, Hey N, Stephenson T, Hill A (1989) *c-myc* oncogene product $p62^{c\text{-}myc}$ in ovarian mucinous neoplasms: immunohistochemical study correlated with malignancy. *J Clin Pathol* 42: 148–152

Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE, Tibshirani R, Botstein D, Borresen-Dale AL, Brown PO (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 99: 12963–12968

Sartelet S, Grossi L, Pasquier D, Conbaret V, Bouvier R, Ranchere D, Plantaz D, Munzer M, Philip T, Birembaut P, Zahm J, Bergeron C, Gaillard D, Pasquier B (2002) Detection of N-myc amplification by FISH in immature areas of fixed neuroblastomas: more efficient than Southern blot/PCR. *J Pathol* 198: 83–91

Sasano H, Nagura H, Silverberg S (1992) Immunolocalization of *c-myc* oncoprotein in mucinous and serous adenocarcinomas of the ovary. *Hum Pathol* 23: 491–495

Sato H, Tsuchiya A, Abe R (1995) Correlation between *c-myc* protein expression and phases of the cell cycle in human colorectal carcinomas. *Fukushima J Med Sci* 41: 111–113

Schraml P, Kononen J, Bubendorf L, Moch H, Bissig H, Noctito A, Mihatsch M, Kallioniemi O, Sauter G (1999) Tissue microarrays for gene amplification surveys in many different tumor types. *Clin Cancer Res* 5: 1966–1975

Sikora K, Chan S, Evan G, Gabra H, Markham N, Stewart J, Watson J (1987) *c-myc* oncogene expression in colorectal cancer. *Cancer* 59: 1289–1295

Sikora K, Evan G, Stewart J, Watson J (1985) Detection of the *c-myc* oncogene product in testicular cancer. *Br J Cancer* 52: 171–176

Smith D, Goh H-S (1996) Overexpression of *c-myc* proto-oncogene in colorectal carcinoma is associated with a reduced mortality that is abrogated by point mutation of the p53 tumor suppressor gene. *Clin Cancer Res* 2: 1049–1053

Sundaresan V, Forgacs I, Wight D, Wilson B, Evan G, Watson J (1987) Abnormal distribution of *c-myc* oncogene product in familial adenomatous polyposis. *J Clin Pathol* 40: 1274–1281

Visca P, Alo P, Del Nonno F, Bott C, Trombetta G, Marandino F, Filippi S, Di Tondo U, Donnorso R (1999) Immunohistochemical expression of fatty acid synthase, apoptotic-regulating genes, proliferating factors, and *ras*, protein product in colorectal adenomas, carcinomas, and adjacent nonneoplastic mucosa. *Clin Cancer Res* 5: 4111–4118

Visscher D, Wallis T, Awussah S, Mohamed A, Crissman J (1997) Evaluation of MYC and chromosome 8 copy number in breast carcinoma by interphase cytogenetics. *Genes, Chromosomes, Cancer* 6: 389–404

Voravud N, Foster C, Gilbertson J, Sikora K, Waxman J (1989) Oncogene expression in cholangiocarcinoma and in normal hepatic development. *Hum Pathol* 20: 1163–1168

Watson J, Stewart J, Evan G, Ritson A, Sikora K (1986) The clinical significance of low cytometric *c-myc* oncoprotein in testicular cancer. *Br J Cancer* 53: 331–337

Williams A, Piris J, Wyllie A (1990) Immunohistochemical demonstration of altered intracellular localization of the *c-myc* oncogene product in human colorectal neoplasms. *J Pathol* 160: 287–293

Yuen M, Wu P, Lau J, Lai C (2001) Expression of c-Myc, c-Fos, and c-Jun in hepatocellular carcinoma. *Cancer* 91: 106–112

Zar JH (1974) *Biostatistical Analysis*. Prentice-Hall: Englewood Cliffs, NJ

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Botstein, et al.
Appl. No.	:	9/866,034
Filed	:	May 25, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Spector, L.
Group Art Unit	:	1647

DECLARATION OF VICTORIA SMITH, Ph.D., UNDER 37 CFR §1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, Dr. Victoria Smith, declare and state as follows:

1. I am a Senior Scientist in the Department of Molecular Biology of Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080.

2. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

3. I joined Genentech in 1996. For approximately three years, I directed a laboratory in the Department of Molecular Biology. During this time I was involved in target discovery for the Tumor Antigen Project, using DNA microarrays to discover genes differentially expressed in tumors compared to their expression in normal tissues. In connection with the above-identified patent application, I directed the generation and analysis of the microarray data attached as Exhibit B.

4. Exhibit B reports the results of the microarray analysis conducted on the gene encoding PRO1800 (DNA35672) as part of the investigation of several newly discovered DNA sequences. The column "Unq Id" identify the gene as 851, which is DNA35672, while the column "DNA Id" identifies the particular lot of PCR product used. The microarray experiments were performed using well-established and accepted microarray techniques known in the art. (*See, e.g.*, Nature Revs. Genetics, 5:229-237 (2004), attached as Exhibit C). The DNA samples used in the microarray studies were obtained from individual lung tumor tissue samples or individual normal lung tissue samples. The individual tumor and normal lung samples were each

compared to pooled samples of normal epithelial tissue. The level of expression in the lung tumor or normal lung tissue sample was compared to the normal pooled epithelial sample, and reported as a raw ratio. The average of the normal lung samples was then used to normalize the data to generate a ratio of expression of the PRO1800 gene in lung tumor samples compared to the average expression in normal lung tissue. In the results reported in Exhibit B, a ratio of 2.0 or greater is a significant result, and indicates a significant increase in expression of the PRO1800 gene in lung tumor tissue compared to the normal lung tissue controls.

5. The results of the microarray studies reported in Exhibit B indicate that the gene encoding PRO1800 (DNA35672) is significantly overexpressed in nine of the eighty lung tumor samples tested compared to the normal lung tissue controls. That is the equivalent of one in every nine samples. In contrast, none of the individual normal lung tissue samples show significant overexpression of the PRO1800 gene. In addition, the average ratio of the lung tumor samples is significantly different from the average ratio of the individual normal lung tumor samples ($p < 0.01$).

6. It is well-established in the art that overexpression of the mRNA for a gene is likely to lead to overexpression of the corresponding protein. Support for this statement can be found, for example, in the Molecular Biology of the Cell, a leading textbook in the field. (Bruce Alberts, *et al.*, Molecular Biology of the Cell (4th ed. 2002), excerpts submitted herewith as Exhibit D). Figure 6-3 on page 302 illustrates the basic principle that there is a correlation between increased gene expression and increased protein expression. The accompanying text states that "a cell can change (or regulate) the expression of each of its genes according to the needs of the moment – *most obviously by controlling the production of its mRNA*." Molecular Biology of the Cell at 302, emphasis added. Similarly, figure 6-90 on page 364 illustrates the path from gene to protein. The accompanying text states that while potentially each step can be regulated by the cell, "the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes." Molecular Biology of the Cell at 364. This point is repeated on page 379, where the authors state that of all the possible points for regulating protein expression, "[f]or most genes transcriptional controls are paramount." Molecular Biology of the Cell at 379.

7. While not every lung tumor sample tested shows overexpression of the PRO1800 gene, the data in Exhibit B indicate that a <u>significant</u> portion of lung tumors do (one in every nine), while none of the normal lung tissue samples show overexpression. Given the known correlation between overexpression of a gene and the corresponding overexpression of the encoded protein, it is very likely that a similar number of lung tumors will overexpress the PRO1800 protein, while normal lung tissue samples will not. Together with the data reported in Example 16 that the gene encoding PRO1800 is amplified in some lung tumors, the results reported in Exhibit B indicate that the PRO1800 gene and protein, as well as antibodies to the encoded protein, can be used to differentiate some cancerous lung tissue from normal lung tissue. Because not all lung tumors show overexpression of PRO1800, it cannot be used to exclude a sample being tested as non-cancerous. However, the PRO1800 gene, protein, and corresponding antibodies are useful as a diagnostic tool since a significant portion of lung tumors overexpress the gene and most likely the encoded protein, while no normal lung samples do.

Appl. No. : **9/866,034**
Filed : **May 25, 2001**

8. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: 01/20/05
Victoria Smith, Ph.D.

S:\DOCS\BSG\BSG-1773.DOC
011905

EXHIBIT A

VICTORIA SMITH

Genentech Inc.
Dept. Molecular Biology
1 DNA Way
South San Francisco CA 94080
Ph: (650) 225 7382
Fax: (650) 225 6497
Email: victoria@gene.com

EDUCATION

Ph.D. (1991) Molecular Biology, University of Cambridge, Cambridge, United Kingdom.
Honors (1987) Biochemistry, University of Western Australia, Australia.
Bachelor of Science (1986) Physical and Inorganic Chemistry, and Biochemistry, University of Western Australia, Australia.

WORK AND RESEARCH EXPERIENCE

Senior Scientist, Genentech Inc (August 1996 – present; promoted to Senior Scientist March 2001)
Lab head, Dept. Molecular Biology

- Identification of potential therapeutic targets for cancer using novel microarray technology
- Discovery and identification of novel secreted proteins
- Development of cancer therapuetics

Stanford University, California, U.S.A. (February 1995 – August 1996)
Research Fellow, Department of Biochemistry
Research: Genomic functional analysis of chromosome V of *Saccharomyces cerevisiae*

Stanford University, California, U.S.A. (January 1992 - January 1995)
Postdoctoral Fellow, Department of Genetics.

Research: Development of new methodology for whole genome functional analysis in microorganisms using genomic sequence data and insertional mutagenesis.

Awards

Human Frontiers Science Program Organization Long Term Fellowship (accepted, 4/01/93 - 1/31/95).

American Cancer Society (California Division) Fellowship (1993, declined).

Cambridge University, United Kingdom (October 1988 - December 1991)
Research undertaken at the Medical Research Council Laboratory of Molecular Biology, Cambridge, UK, for the degree of Doctor of Philosophy, Cambridge University.
Thesis: A Molecular Genetic Analysis of Yeast Chromosome IX.
Thesis Advisor: Dr. Barclay Barrell.

Awards

Max Perutz Prize in 1991 for outstanding performance by a graduate student. Awarded for advances in genomic-scale DNA sequencing methodology, and genetic analysis of the *SNP1* gene of *Saccharomyces cerevisiae* .

King Edward Memorial Hospital for Women, Western Australia (1988)
Research Scientist.
Research: Analysis of hormone-inducible mRNAs in breast tumors.

University of Western Australia (1984 - 1987)
1984 - 1986: Bachelor of Science degree with double major in Biochemistry and Physical and Inorganic Chemistry.
1987: First Class Honors in Biochemistry. Thesis: Nuclease Sensitivity and Methylation Patterns of the Phenylalanine Hydroxylase Gene.

Awards

Association of Commonwealth Universities Scholarship for study in the United Kingdom (accepted, October 1988 - October 1991, University of Cambridge).
Hackett Scholarship for overseas study (1988, declined).
Lady James Prize in Natural Science in 1986: Best student completing Bachelor of Science degree with a major in a natural science.
JWH Lugg Prize in Biochemistry in 1986: Best student completing major in Biochemistry.

Convocation Prize in Science in 1985: Best student completing major in second year Physics, Geology or Chemistry.
Shell Prize in Chemistry in 1985: Best student completing major in second year Chemistry.

PUBLICATIONS

V. Smith, E.F. Shen, D. Wieand, T.H. Landon, N.A. Wong, A.M. Lessells, S. Paterson-Brown, T.D. Wu, J.Z. Tang, K.J. Hillan, I.D. Penman, Expression Analysis of the Metaplasia-Dysplasia-Carcinoma Sequence in Barrett's Esophagus (submitted).

Tice DA. Szeto W. Soloviev I. Rubinfeld B. Fong SE. Dugger DL. Winer J. Williams PM. Wieand D. Smith V. Schwall RH. Pennica D. Polakis P. *Journal of Biological Chemistry.* 277(16):14329-35, 2002 Apr 19.

N.J. Maughan, F. Lewis, V. Smith (2001) *Journal of Pathology* 195, 3-6.

F. Lewis, N.J. Maughan, V. Smith, K. Hillan, P. Quirke (2001) *Journal of Pathology* 195, 66-71.

D.J. Garfinkel, M.J. Curcio and V. Smith (1998) Ty Mutagenesis *Methods in Microbiology,* volume 26, 101-117.

C. Churcher, S. Bowman, K. Badcock, A. Bankier, D. Brown, T. Chillingworth, R. Connor, K. Devlin, S. Gentles, N. Hamlin, D. Harris, T. Horsnell, S. Hunt, K. Jagels, M. Jones, G. Lye, S. Moule, C. Odell, D. Pearson, M. Rajandream, P. Rice, N. Rowley, J. Skelton, V. Smith, S. Walsh, S. Whitehead & B. Barrell. (1997) *Nature,* 387, 84-87.

F. S. Dietrich, J. Mulligan, K. Hennessy, M. A. Yelton, E. Allen, R. Araujo, E. Aviles, A. Berno, T. Brennan, J. Carpenter, E. Chen, J. M. Cherry, E. Chung, M. Duncan, E. Guzman, G. Hartzell, S. Hunicke-Smith, R. W. Hyman, A. Kayser, C. Komp, D. Lashkari, H. Lew, D. Lin, D. Mosedale, K. Nakahara, A. Namath, R. Norgren, P. Oefner, C. Oh, F.X. Petel, D. Roberts, P. Sehl, S. Schramm, T. Shogren, V. Smith, P. Taylor, Y. Wei, D. Botstein & R. W. Davis. (1997) *Nature* 387, 78-81.

V. Smith, K. Chou, D. Lashkari, D. Botstein, and P. O. Brown. (1996). Functional Analysis of the Genes of Yeast Chromosome V by Genetic Footprinting. *Science* 274, 2069-74

V. Smith, D. Botstein, and P. O. Brown (1995) Genetic Footprinting: A genomic strategy for determining a gene's function given its sequence. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 6479-6483.

V. Smith, M. Craxton, A. T. Bankier, C. M. Brown, W. D. Rawlinson, M. Chee, and B. G. Barrell (1995) Microtiter methods for the preparation and fluorescent sequencing of M13 clones. *Recombinant DNA Methodology II: a volume in the Selected Methods in Enzymology Series,* pp. 607 - 621.

V. Smith, M. Craxton, A. T. Bankier, C. M. Brown, W. D. Rawlinson, M. S. Chee, and B. G. Barrell (1993) Microtiter methods for the preparation and fluorescent sequencing of M13 clones. *Methods in Enzymology,* 218 , 173-187.

V. Smith and M. S. Chee (1991) A simple method for sequencing the complementary strand of ssDNA from M13 clones. *Nucleic Acids Research* 19, 6957.

V. Smith and B. G. Barrell (1991) Cloning of a Yeast U1 snRNP 70K protein homologue: functional conservation of an RNA binding domain between humans and yeast. *EMBO Journal* 10, 2627-2643.

W. D. Rawlinson, M. S. Chee, V. Smith and B. G. Barrell (1991) Preparation of large numbers of single stranded DNA templates by rescue from phagemids in microtiter plates. *Nucleic Acids Research* 19, 4779.

V. Smith, C. M. Brown, A. T. Bankier, and B. G. Barrell (1990) Semi-automated preparation of DNA templates for large scale sequencing projects. *DNA Sequence* 1, 73-78.

S. J. Wysocki, E. Hahnel, S. P. Wilkinson, V. Smith, and R. Hahnel (1990) Hormone-sensitive gene expression in breast tumors. *Anticancer Research* 10, 185-188

S. J. Wysocki, E. Hahnel, A. Masters, V. Smith, A. J. McCartney and R. Hahnel (1990) Detection of pS2 messenger RNA in gynecological cancers. *Cancer Research* 50, 1800-1802

PATENTS

Genetic Footprinting: Insertional Mutagenesis and Genetic Selection. U.S. Patent No. 5,612,180. Inventors: Patrick Brown and Victoria Smith

PATENTS FILED (at Genentech Inc.)

Methods of Detecting and Quantifying Gene Expression. Inventors: Victoria Smith, Edward Robbie, David Lowe, James Marsters.

Compositions and Methods for the Treatment of Cancer. Inventors: Victoria Smith, Austin Gurney, Audrey Goddard, Fred DeSauvage

Diagnostic for Dysplasia in Barrett's Esophagus. Inventor: Victoria Smith

Numerous composition of matter filings related to novel gene discovery, pending

EXHIBIT B

Unq Id	DNA Id	Experiment Name	Raw Ratio (sample/pooled epithelial)	Normalized Ratio (sample/normal lung)
Lung Tumor Samples				
851.	157,736.	100ngLungBaCa1069 vs 25ngEpi1409	0.563	0.990
851.	157,736.	lung 1431 vs epi pool	0.634	1.115
851.	157,736.	lung SqCa R694 vs epi pool	0.847	1.489
851.	157,736.	Lung SqCa-hf 1649	0.337	0.593
851.	157,736.	Lung SqCa-hf 1649	0.209	0.367
851.	157,736.	lung tumor 1055 vs epi pool	1.085	1.908
851.	157,736.	lung tumor 10ng	0.858	1.509
851.	157,736.	lung tumor 1370 vs epi pool	0.404	0.710
851.	135,920.	lung tumor 1647	0.444	0.781
851.	135,920.	lung tumor 1648	0.633	1.113
851.	157,736.	lung tumor 1685/ref.RNA	0.886	1.558
851.	135,920.	Lung Tumor 685	0.18	0.316
851.	157,736.	lung tumor 688 vs epi pool	0.443	0.779
851.	135,920.	lung tumor 734	0.456	0.802
851.	135,920.	lung tumor 735	2.136	3.756
851.	135,920.	lung tumor 737	0.789	1.387
851.	135,920.	lung tumor 738	0.884	1.554
851.	135,920.	lung tumor 739	0.655	1.152
851.	101,241.	Lung tumor hf 842	0.849	1.493
851.	135,920.	Lung Tumor HF-0017083	0.467	0.821
851.	135,920.	lung tumor hf-1291	0.078	0.137
851.	157,736.	lung tumor hf-1333	0.496	0.872
851.	157,736.	lung tumor hf-1340	0.514	0.904
851.	157,736.	lung tumor hf-1364	0.343	0.603
851.	157,736.	lung tumor hf-1366	0.634	1.115
851.	157,736.	lung tumor hf-1587	0.478	0.840
851.	157,736.	lung tumor hf-1587 v normal	0.442	0.777
851.	157,736.	lung tumor hf-1587 v. normal	0.404	0.710
851.	135,920.	lung tumor hf-1596	0.491	0.863
851.	135,920.	lung tumor hf-1646	0.258	0.454
851.	135,920.	lung tumor hf-1649	0.278	0.489
851.	135,920.	lung tumor hf-1655	0.597	1.050
851.	135,920.	lung tumor hf-1719	0.612	1.076
851.	157,736.	lung tumor hf-1775	0.422	0.742
851.	157,736.	lung tumor hf-1785	0.429	0.754
851.	135,920.	Lung Tumor HF1602	0.346	0.608
851.	135,920.	Lung Tumor HF1651	0.592	1.041
851.	135,920.	Lung Tumor HF1729	0.611	1.074
851.	101,241.	Lung Tumor HF631	0.627	1.102
851.	101,241.	Lung tumor hf840	0.579	1.018
851.	157,736.	lung tumor R1057 vs epi pool	0.523	0.920
851.	157,736.	lung tumor R1094 vs epi pool	1.004	1.765
851.	157,736.	lung tumor R1094 vs epi pool	0.813	1.430
851.	157,736.	lung tumor R1372 vs epi pool	0.808	1.421
851.	157,736.	lung tumor R1372 vs epi pool	0.739	1.299
851.	157,736.	lung tumor R417 vs epi pool	0.431	0.758

851.	157,736. lung tumor R542 vs epi pool	0.454	0.798
851.	157,736. lung tumor R543 vs epi pool	0.651	1.145
851.	157,736. lung tumor R544 vs epi pool	1.407	2.474
851.	157,736. lung tumor R544 vs epi pool	0.655	1.152
851.	157,736. lung tumor R685 vs epi pool	0.894	1.572
851.	157,736. lung tumor R693 vs epi pool	0.836	1.470
851.	157,736. lung tumor R693 vs epi pool	0.527	0.927
851.	157,736. lung tumor R737 vs epi pool	0.479	0.842
851.	157,736. lung tumor R742 vs epi pool	0.526	0.925
851.	157,736. lung tumor R777 vs epi pool	1.041	1.830
851.	157,736. lung tumor R777 vs epi pool	0.573	1.008
851.	157,736. lung tumor R789 vs epi pool	0.849	1.493
851.	157,736. lung tumor R789 vs epi pool	0.707	1.243
851.	157,736. lung tumor R791 vs epi pool	0.869	1.528
851.	157,736. lung tumor R791 vs epi pool	0.711	1.250
851.	135,920. Lung Tumor RNA689	0.209	0.367
851.	135,920. Lung Tumor RNA691	0.536	0.942
851.	135,920. Lung Tumor RNA693	0.441	0.775
851.	157,736. lung tumor vs epi pool	16.613	29.211
851.	157,736. lung tumor vs epi pool	0.443	0.779
851.	157,736. lung tumor	15.387	27.055
851.	157,736. lung tumor/Ref.RNA	0.761	1.338
851.	157,736. lung tumor1582/Ref.RNA	0.92	1.618
851.	157,736. lung tumor1683/epipool	1.35	2.374
851.	157,736. lung tumor1685/epipool	0.707	1.243
851.	157,736. lung tumor1853/REF.RNA	1.491	2.622
851.	101,241. lung tumorhf641	0.911	1.602
851.	157,736. LungBaCa vs Epi	0.342	0.601
851.	157,736. LungBAca1662 vs Epi	3.362	5.911
851.	157,736. LungBAca	2.466	4.336
851.	157,736. LungCaSq-hf1602	0.713	1.254
851.	157,736. LungCaSq-hf1647	0.885	1.556
851.	157,736. LungSqCa-hf1293	1.291	2.270
851.	157,736. LungSqCa-hf1646	0.411	0.723
851.	157,736. LungSqCa-hf?	1.112	1.955

Normal Lung Samples

851.	157,736. normal lung 795	0.221	0.389
851.	157,736. normal lung hf-1773	0.516	0.907
851.	157,736. N. lung R1415 vs univ. ref	0.991	1.742
851.	157,736. N. lung R1431 vs univ.ref	0.709	1.247
851.	157,736. N. lung R417 vs univ.ref	0.669	1.176
851.	157,736. lung Normal 1052 vs epi pool	0.603	1.060
851.	157,736. lung Normal 1417 vs epi pool	0.317	0.557
851.	157,736. lung Normal R417 vs epi pool	0.982	1.727
851.	157,736. lung Normal R419 vs epi pool	0.335	0.589
851.	157,736. 423LungInflamTA1	0.517	0.909
851.	157,736. 488-551LungInflamTA1	0.486	0.855
851.	157,736. 488-552InflmdLungTA1	0.425	0.747
851.	157,736. 1054LungInflamTA1	0.56	0.985

851.	157,736.	1415LungInflamTA1	0.706	1.241
851.	157,736.	1415LungInflamTA1	0.494	0.869

EXHIBIT C

GUIDELINES

Expression profiling — best practices for data generation and interpretation in clinical trials

*The Tumor Analysis Best Practices Working Group**

Microarrays are routinely used to assess mRNA transcript levels on a genome-wide scale. As use and acceptance increases, there is intensified focus on appropriate methods of data generation and interpretation, with important questions being asked about the best data analysis methods. The development of such 'best practices' is needed, as microarrays — in particular, Affymetrix oligonucleotide arrays — are becoming increasingly important in human clinical trials, both for differential diagnosis and monitoring of pharmacological efficacy. Here, representatives from high-volume microarray core centres consider the current status of 'best practices', focusing on the broadly used Affymetrix oligonucleotide arrays.

Microarrays represent a major technological advance in molecular biology. The introduction of any such advance is typically followed by a period of optimization and standardization. The latter is a crucial part of any maturing technology, as it allows an approach in which advances are made in parallel by individual researchers and companies who contribute new knowledge based on the existing standard. Any such standards must be constantly reassessed; stale or stagnant standards can inhibit the development of the technology.

Microarray-based mRNA-expression profiling can be considered to be the first mature genome-wide analysis technology, reflected in an increased interest in using microarrays as an endpoint in clinical trials. However, regulations of clinical trials require the development of clear standards for use and interpretation of microarray data (commonly referred to as quality control and standard operating procedures (QC/SOPs) and/or 'best practices'). Guidelines for reporting and annotation of microarray data from the Microarray Gene Expression Data (MGED) Society (see online links box) — using MIAME (Minimum Information About A Microarray Experiment) standards (BOX 1) and the MAGE-ML mark-up language[1,2] — represent an important step towards this goal. The efforts of this multinational academic–industry partnership has made it possible to develop databases that can house the many types of microarray data (see below) within the same data structure, enabling some data queries between experiments and experimental platforms. The ArrayExpress microarray database[3] (see online links box) is the first major publicly accessible database that adheres to this universal data-presentation platform, and some prominent journals (such as *Nature*, *Cell*, *EMBO Journal* and *The Lancet*) now demand that published microarray data conform to the MIAME standards. In addition, microarray manufacturers, such as Affymetrix, have implemented MIAME-compliant data output in their new software releases.

The MGED Society has effectively developed data-reporting guidelines, but it has not addressed issues of data generation and interpretation. The latter are more intimately coupled to the specific experimental platform. Of the three commonly used types of microarrays (spotted cDNA, spotted oligonucleotide and Affymetrix arrays), each has distinct methodologies associated with them; accordingly, the issues of data interpretation are also different (BOX 2). These differences make it difficult or impossible to develop cross-platform guidelines for data generation and interpretation. Best practices for spotted cDNA arrays are especially problematic because the manufacture of the arrays varies considerably from place to place. In addition, all spotted arrays use co-hybridization of a test RNA sample labelled with one colour FLUOROPHORE with a control RNA labelled with a different colour to which the test is compared on the same spot. The output is in the form of a ratio of hybridization signals that is comparable to other experiments only if the same control RNA is always used. Therefore, the development of standards in spotted arrays would require all laboratories to use the same control RNA solution before data could be easily compared.

Manufactured oligonucleotide arrays (both mechanically spotted and synthesized *in situ*) have the advantages of being centrally produced under controlled conditions. Affymetrix PHOTOLITHOGRAPHY-produced arrays have been available for nearly 10 years, whereas mechanically spotted oligonucleotide arrays have only very recently begun to appear in the marketplace. For example, Agilent Technologies (see online links box) recently released 17,000 60-mer oligonucleotides printed five times each on glass slides (85,000 FEATURES). Spotted oligonucleotide arrays typically have a single spot per gene (single probe measurement), whereas Affymetrix arrays provide multiple measurements — a series of independent or semi-independent oligonucleotides query each RNA in solution (the probe set) (BOX 2). Affymetrix probe sets are constructed from a series of perfect-match and paired-mismatch oligonucleotides, allowing some assessment of non-specific binding and performance of the probes. Overall, the Affymetrix probe sets provide a variety of measurements that allow robust measures of gene expression. The use of multiple perfect-match and mismatch probes for each gene enables the development of different methods of interpreting the hybridization patterns across the probe set and calculating a single 'expression level' or 'signal' that reflect the gene's relative expression level. A number of probe-set interpretation algorithms for Affymetrix arrays are available (see below for discussion).

> "[distinct methodologies] make it difficult or impossible to develop cross-platform guidelines for data generation and interpretation."

The increasing use of Affymetrix microarrays, and the emergence of this technology as an endpoint in clinical trials, has led to requests to develop, in both the pharmaceutical and academic research communities, best practices in data generation and analysis. Given the many differences between spotted cDNA, spotted oligonucleotide and Affymetrix arrays, the best practices need to be developed separately for each experimental platform; this is in contrast to data reporting that can be standardized across all platforms (BOXES 1 and 2). The Tumor Analysis Best

Box 1 | The MIAME guidelines for data reporting

The Microarray Gene Expression Data Society (MGED) is an international discussion group of microarray experts, with the primary goal of developing methods for data sharing between experimental platforms. The main output of this group has been the Minimum Information About A Microarray Experiment (MIAME) guidelines for microarray data annotation and reporting. The guidelines have been adopted by a number of scientific journals and have recently been endorsed for use by the US Food and Drug Administration and the US Department of Agriculture for pharmacogenomics projects.

The MIAME guidelines include descriptions of experimental design (number of replicates, nature of biological variables), samples used, extract preparation and labelling, hybridization procedures and parameters, and measurement data and specifications. These guidelines have been most important for the spotted cDNA and oligonucleotide experimental platforms (see BOX 2) in which the flexibility in microarray design and utilization also leads to considerable variation in array data generation and reporting between different laboratories. The guidelines do not attempt to dictate how experiments should be done, but rather provide adequate information associated with any published or publicly available experiment so that the experiment can be reproduced.

Box 2 | Microarray experimental platforms

There are three different types of microarray in common use: spotted cDNAs, spotted oligonucleotides and Affymetrix arrays.

Spotted cDNA arrays

Spotted cDNA arrays typically use sets of plasmids of specific cDNAs in gridded liquid aliquots. The inserts of each clone are typically amplified by PCR, and a few picolitres are physically spotted onto glass slides by liquid-handling robots. Robotic spotters can spot 100,000 spots per slide, and duplicate sets of clones are often spotted. The advantages of spotted cDNA arrays are that the content of each microarray is determined by the researcher, with complete flexibility in number and type of cDNA clones spotted. Also, the cost per array is relatively low, as the clone sets are a PCR-renewable resource and the glass slides are themselves inexpensive. The amount of the RNA that corresponds to each spot is determined relative to a second control RNA solution that is hybridized to the same spot, and a ratio is obtained.

Disadvantages of spotted cDNA arrays include the variable amount of DNA spotted in each spot, the 10–20% 'drop out' rate of failed PCR reactions or failed spots and mis-identification of clones (that is, the spot is not what you think it is). Also, there is no control over the actual sequence of the clone. As many gene-coding sequences contain regions of sequence that are shared with other genes, there are questions of specificity of the hybridization to the relatively large cDNA inserts. Spotted cDNA arrays were embraced by most academic centres, owing to their flexibility and relatively low cost.

Spotted oligonucleotide arrays

These arrays are also built by liquid handling on glass slides; however, the input solution is a synthetic oligonucleotide (often 60–70-mers). The resulting spotted material is typically of known concentration, of known sequence and is single stranded (all advantages relative to spotted cDNAs). Most of the process can be automated, leading to less sample mix-up and less drop-out of samples.

Disadvantages of spotted oligonucleotides include the relatively high cost of synthesizing large numbers of large oligonucleotides and the non-renewable nature of the resource. Spotted oligonucleotide arrays are becoming increasingly available.

Affymetrix GeneChips

These microarrays are factory designed and synthesized. Design is done using software to choose a series of 11 25-mer probes from the 3′ end of each transcript or predicted transcript in the genome; each of the 11 probes is then paired with a similar mismatch probe that is designed to contain a mutation in the centre. The latter serves as a form of control for hybridization specificity. Synthesis of arrays is done using light-activated chemistry and photolithography methods, and feature size can be reduced to approximately 8 μm^2, with about 1 million probes in a 1.2 cm^2 glass area. Probe-set algorithms interpret the signals from each 22-oligonucleotide probe set, and derive a single value (signal) from the patterns of hybridization to the 22 individual probes. This signal is then normalized to the entire microarray, or to the probe sets across an entire project.

For a more general discussion of normalization and analysis methods of different microarray platforms, the reader is referred to the excellent web information resource of the MGED group (see The MGED Data Transformation and Normalization Working Group in online links box).

Practices Working Group (see BOX 3) was convened to discuss and develop best practices for Affymetrix microarrays, including QC and SOPs for both data generation and data analyses. The first meeting was held in Santa Clara in March 2003, followed by a series of conference calls that focused on discussions of data generation and analysis standards for the Affymetrix oligonucleotide arrays. The Working Group deliberately focused on a platform that has widespread usage and is most likely to be used in clinical trials owing to the previously standardized manufacturing process. Here, we discuss recommendations for experimental design, probe-set analysis algorithms, signal/noise assessments and biostatistical methods.

Experimental design

Appropriate experimental design is a key aspect of all science, and microarray studies are no exception. The relatively high cost of some commercial microarray platforms is a frequently cited reason for suboptimal experimental design, especially with regards to the number of replicates. Data interpretation is inevitably compromised when replicates are decreased.

Replication in cross-sectional studies. The appropriate number of microarray replicates for any particular condition or time point depends on the source of biological variability in the study samples. Inter-individual variability is very large in outbred (genetically heterogeneous) humans, but is very small within inbred mouse strains. For example, expression profiles derived from muscle from different mice are not more variable than from muscles isolated from one mouse[4]. Defining the confounding variables that contribute to experimental variability, such as intra-subject, inter-subject, inter-group and technical variation (microarray protocol), is needed to design and statistically power a study, and to determine the number of replicates that are needed. In general, inbred mice require testing only three or four mice per group. We and others have found that five or six out-bred rats per group provide statistically robust results[5,6]. By contrast, human samples require considerably more individuals per group. Key variables in human samples include tissue heterogeneity, stage of disease and inter-individual variation, all of which have been found to be major confounding variables[7].

Replication in longitudinal studies. It has long been recognized that, in human clinical trials, LONGITUDINAL DESIGNS provide considerably greater power at lower numbers of replicates.

They best control for inter-individual variation because each subject serves as their own control. Serial blood sampling from single subjects is the least invasive[8] (see below for further discussion), and, for example, cancer patients are often longitudinally sampled[9]. Serial biopsies of other tissues are more invasive: however, a number of serial human muscle biopsy studies of healthy volunteers after different types of exercise training have begun to appear in the literature[10,11].

Expression profiling of blood samples (longitudinal or CROSS-SECTIONAL DESIGN) is the protocol that is most likely to be used in human clinical trials. One of the Working Group's goals was to establish SOPs for blood sample collection and RNA isolation in clinical trials. A specific follow-up report of these recommendations will be published elsewhere. Such a protocol must be easily adaptable to multiple trial sites, with relatively little need for resident expertise to carry out the isolation protocol. So far, standard methods for isolating peripheral-blood mononuclear cells have shown the most reproducibility, although others are being tested (see Affymetrix Technical Note in online links box). Cells isolated soon after collection can be flash frozen for storage and subsequent RNA isolation or an RNA stabilizing compound can be added if the samples need to be transported.

Tissue/cell heterogeneity. Tissue heterogeneity is a major confounding variable in most microarray experiments. In inbred mice, tissue heterogeneity is typically normalized by using whole organs. This is rarely possible in human experiments, and particularly not in clinical trials; the limited amount of human tissue that is available exacerbates heterogeneity. The mixed cell populations of peripheral blood can be thought of as a tissue heterogeneity problem similar to that encountered in all solid tissue and tumour biopsies. Indeed, a recent study showed that variation as a result of tissue variability in human muscle biopsies often exceeded inter-individual variability[12]. One potential solution to the tissue heterogeneity problem lies in bioinformatic methods. If computer software can be trained to recognize the expression profiles of each individual cell type within a mixed tissue sample, then it should be possible to subtract them from each other and to renormalize to obtain a set of cell-specific expression profiles derived from a single mixed profile. This will be most easily done on tumour biopsies, in which the main cells of interest are tumour versus contaminating normal tissue. Although there are no published examples so far, such methods are maturing rapidly.

Box 3 | The Tumor Analysis Best Practices Working Group*

The Tumor Analysis Best Practices Working Group is a group of investigators who study the best practices of tumour analysis in humans taking part in clinical trials. The following authors are members of the Group:

- Eric P. Hoffman is at the Research Center for Genetic Medicine, Children's National Medical Center, Washington DC 20010, USA. email: ehoffman@cnmcresearch.org
- Tarif Awad, John Palma, Teresa Webster, Earl Hubbell and Janet A. Warrington are at Affymetrix, Santa Clara, California 95051, USA. emails: tarif_awad@affymetrix.com; john_palma@affymetrix.com; teresa_webster@affymetrix.com; earl_hubbell@affymetrix.com; janet_warrington@affymetrix.com
- Avrum Spira is at The Pulmonary Center, Boston University Medical Center and the Bioinformatics Program, Boston University, Boston, Massachusetts 02118, USA. e-mail: aspira@lung.bumc.bu.edu
- George Wright is at the Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institute of Health, Bethesda, Maryland 20892, USA. e-mail: wrightge@mail.nih.gov
- Jonathan Buckley and Tim Triche are at the Children's Hospital, University of California, Los Angeles, California 90089, USA. e-mail: buckley@hsc.usc.edu; triche@hsc.usc.edu
- Ron Davis, Robert Tibshirani and Wenzhong Xiao are at Stanford University, Palo Alto, California 94303, USA. e-mails: dbowe@stanford.edu; tibs@stat.stanford.edu; wzxiao@pmgm2.stanford.edu
- Wendell Jones is at Expression Analysis Inc., Durham, North Carolina 27713, USA. e-mail: wjones@expressionanalysis.com
- Ron Tompkins is at Harvard University, Boston, Massachusetts 02115, USA. e-mail: rtompkins@partners.org
- Mike West is at the Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708, USA. e-mail: mw@stat.duke.edu

An experimental alternative to mitigate confounding tissue heterogeneity is to isolate pure cell populations for expression profiling. Many such methods are well developed in the research laboratory, including FLUORESCENCE-ACTIVATED CELL SORTING (FACS)[13], NEGATIVE CELL ISOLATIONS from blood (for example, Stem Cell Technologies RosetteSep)[14] and LASER CAPTURE MICRODISSECTION[15]. To research scientists, the profiles that are derived from isolated cell types are a more intuitive approach to define biologically relevant pathways. However, it should be noted that uses of array-based analysis of gene expression approved by the US Food and Drug Administration (FDA) will probably focus on reproducibility and robustness (as well as on predictive accuracy), rather than on biological interpretation or justification. The high-tech methods used to isolate specific cell types from clinical samples are unlikely to make their way into clinical trials unless tissues are procured in a highly centralized way.

"If computer software can be trained to recognize the expression profiles of each individual cell type within a mixed tissue sample, then it should be possible to subtract them from each other..."

Procedural variation. Beyond the usual issues of sampling and accrual, gene-expression data will be subject to many additional sources of error. For example, the surgical removal and processing of tumour tissue can vary considerably from site to site. Laboratory QC procedures in tissue handling, RNA extraction and processing, and variations in protocols for data management and processing will need to be addressed in any clinical trial design. In particular, prolonged tissue ISCHAEMIA prior to processing of surgically RESECTED tissue can significantly alter gene expression[16]. All tissue samples should be flash frozen within minutes of surgery and stored at –80°C or below. Samples should also be kept in small, airtight containers and kept from drying out during frozen storage by placing fragments of ice in with the sample.

Technical variability

The standard laboratory protocol for generating RNA profiles using Affymetrix microarrays involves a series of steps (FIG. 1).

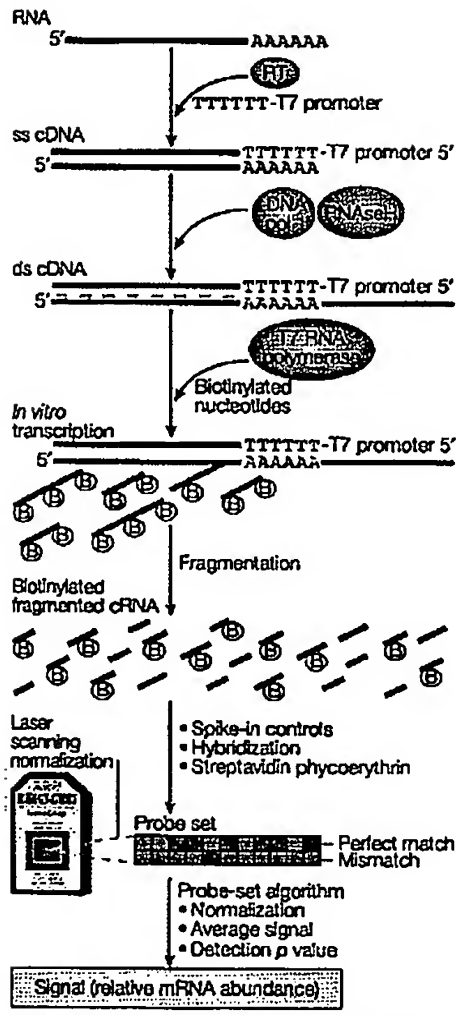


Figure 1 | Sample processing and microarray interpretation of Affymetrix GeneChips. Flash-frozen tissue (~50 mg) is homogenized to isolate total RNA. Single-stranded (ss) cDNA and then double-stranded (ds) cDNA is made from ~5 μg of total RNA. Double-stranded cDNA contains a T7 RNA-polymerase promoter adjacent to the 3′ polyA tail of each transcript. It is transcribed *in vitro* to generate more than 400 biotinylated cRNA molecules for each ds cDNA molecule. The biotinylated cRNA is fragmented and hybridized to the microarrays. Each transcript is queried by one or more probe sets of 11 perfect-match and 11 paired-mismatch oligonucleotides (the latter contain a centrally located point mutation as a form of hybridization specificity control). Currently available Affymetrix microarrays have ~54,000 probe sets on each 1.28 cm^2 glass microarray (~1.2 million 25-mer oligonucleotides on the HG-U133Plus 2.0 array). The biotinylated cRNA fragments hybridize to the appropriate oligonucleotide features. A laser scanner determines the amount of bound biotinylated cRNA indirectly through the streptavidin-conjugated phycoerythrin fluorescence at each feature within a probe set. The component probe pairs are interpreted and averaged to arrive at a single signal that reflects the relative abundance of the original mRNA. Probe sets are interpreted by any one of a number of probe-set algorithms, each providing a signal that reflects the relative hybridization intensity across the probe set. RT, reverse transcriptase.

RNA isolation. RNA quality and quantity is crucial to the success and reproducibility of the expression profiles. RNA quantity and quality is generally checked by complementary methods: UV 260/280 ratio >1.8, agarose gel electrophoresis or an Agilent Bioanalyzer to visualize clear 18S and 28S ribosomal RNA bands. Total RNA (5–10 μg) is input into the cDNA/cRNA reaction, with an expected corrected yield of biotinylated cRNA of between 4- and 10-fold greater than the total RNA input (so 5 μg of total RNA must yield at least 20 μg of biotinylated cRNA, or the sample is discarded). The biotinylated cRNA should be 500–3,000 base pairs (bp) in size. After fragmentation, the cRNA should be 50–200 bp. The Working Group recommends that samples that do not meet these criteria should be discarded.

If RNA amount is limiting — as is the case, for example, with laser capture microscopy samples, flow-sorted cell samples or small tissue samples — a two-round amplification protocol can be used. For example, 200 ng of total RNA is processed for *in vitro* transcription (IVT), with the same goal of 4–10-fold amplification (>800 ng of cRNA output). One hundred nanograms of this cRNA is then reverse transcribed into cDNA using random primers, after which a second IVT is done. The second round IVT should result in a 400-fold amplification.

Microarray controls. Hybridization controls include visualization of the image so that any abnormalities in hybridization patterns can be detected. ProbeProfiler from Corimbia Inc. is a program with extended capabilities for detecting defects in microarray manufacture. Affymetrix MAS 5.0 software adjusts the microarray-scanned image to a common target intensity by using a scaling factor. In addition, a general index of chip background and noise is represented by the percentage of 'present calls' (probe sets for which the hybridization to the perfect-match probes is significantly higher than mismatch hybridization). The Working Group believes that both the scaling factor and the percentage of present calls are important QC criteria. Considering MAS 5.0 chip analyses, the scaling factors to normalize chips within a project should lie within two standard deviations of the mean, with present calls being greater than 25% (BOX 4). The percentage of present calls is often lower when B or C arrays that contain higher proportions of more poorly characterized transcript units (expressed sequence tags or computer-predicted open reading frames) are used. The percentage of present calls across a set of samples should be consistent, within a range of 10%. Some software packages allow the identification of statistical 'outlier' microarrays in a group of microarrays in a given project, which additionally enables the experimenter to flag and exclude specific microarrays that are not acceptable for an analysis. In addition to these criteria, acceptable hybridizations must have adequately intact input RNA as shown by 3′ to 5′ ratios of hybridization within probe sets. A typical control is the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene, which should have 3′ to 5′ ratios of less than 3 (BOX 4).

The QC criteria provided above are based on MAS 5.0 probe-set algorithms and data analyses. The measures of present calls and scaling factors are useful and serve as initial summary measures of the performance of a particular microarray. However, more focused statistical methods, coupled with routine visual inspection of images, hold promise for the continuing improvement of data quality and screening abilities.

Large-scale analyses of microarray data across laboratories have not yet been reported. However, the Working Group feels that adherence to the above QC criteria, using standard RNA isolation and processing methods, should yield data that are consistent between laboratories and intrinsically comparable. The same set of criteria can also be used as best practices for data generation in the design and conduct of clinical trials.

Standard clinical laboratory practice is to develop programmes for submission of known samples to different laboratories and assessment of comparability of results. Such programmes are under development within larger collaborative efforts, such as the National Heart, Lung and Blood Institute (NHLBI) Programs in Genomic Applications (see the HOPGENE Program for Genomic Applications in online links box) and the National Institute of General Medical Sciences (NIGMS) Glue Grant (see online links box).

Data analysis and interpretation

Signal generation versus statistical analyses. Two relatively distinct steps underlie all data analyses of Affymetrix oligonucleotide microarrays: the development of a normalized 'signal' for each transcript on each microarray and the subsequent statistical analysis of differences in signals between different arrays. The first step involves probe-set algorithms that use all, or part, of the component signals within a probe set and then derive a single signal that is representative of the relative abundance of each mRNA queried in each array. The second step is the

application of bioinformatic and statistical methods to identify interesting subsets of the assembled data of all arrays within a project. There is considerable debate about the best methods for both of these steps (see below for a discussion). Although the two steps are separable, it is clear that they have a marked influence on each other. It is in this realm that the bioinformatics of microarrays becomes avant-garde, and with the groundbreaking nature of research comes considerable debate as to what is appropriate in any specific situation.

Before discussing the different methods for probe-set analysis and data interpretation, it is important to point out that much of the debate in the field of bioinformatics about microarray interpretation revolves around signal/noise ratios. A common assumption is that signal/noise ratios across a microarray are homogeneous, or at least similar in magnitude. This might be true for general background hybridization, but not for the performance of probe sets. In any particular microarray, there are probe sets that give very strong and clear hybridization patterns and those that perform poorly. Many of the best performing probe sets (those with a highly significant probe-set detection *p* value) reflect highly expressed transcripts with no closely related sequences that might cross-hybridize. Low-level transcripts, or transcripts that belong to gene families with highly homologous sequences derived from distinct genes, often have corresponding probe sets that do not perform as well and might have a significant, if not overwhelming, noise component. The signal from such probe sets is difficult to interpret, and data interpretation can be limited to only the best performing probe sets, although arguably the most interesting data comes from the genes that are expressed at low levels but that still show significant differences between samples.

Determining adequate sensitivity of the signals and signal/noise responses relative to the absolute quantity of mRNA in clinical samples is crucial as microarrays become a component of clinical trials and diagnostic models. Affymetrix arrays provide a concentration of each mRNA queried relative to the genome-wide mRNA profile of the sample; it is assumed that the global mRNA content of a tissue as a whole does not change significantly, making relative mRNA quantification an accurate reflection of the response of the individual gene. This method differs from absolute quantification of specific mRNAs (such as S1 NUCLEASE PROTECTION and REAL-TIME PCR), or the isolated transcript ratio determined by co-hybridization of two samples to spotted cDNA or oligonucleotide arrays (BOX 2).

Box 4 | Quality Control metrics for Affymetrix microarrays

RNA quality

Optical density 260/280 of 1.8–2.1 | Agilent Bio-Analyzer | Gel electrophoresis

cDNA/cRNA efficiency

>4-fold amplification from total RNA | 500–3,000 bp prior to fragmentation | 50–200 bp after fragmentation

Chip hybridization

Image inspection for defects | Scaling factors within two standard deviations within a project | MAS 5.0 present calls >25% for the A-SERIES ARRAYS, including the HG-U133Plus 2.0 array | Percentage present calls for the B- AND C-SERIES ARRAYS are typically lower | 3'/5' *GAPDH* ratios <3

Project normalization

The detection of statistical outliers for chips, probe sets or individual probe pairs requires normalization and analysis across an entire project. This is afforded by the dCHIP and ProbeProfiler, and other software packages. Data-analysis packages that rely on intra-chip normalization and scaling typically do not enable detection of statistical outliers.

Chip outliers

Probe-set outliers | Probe-pair outliers | Range in present calls <10%

Affymetrix arrays achieve considerable sensitivity through the inherent redundancy of the probe set; however, the Working Group acknowledged that some genes, such as some cytokines that are functional at very low expression levels, are probably below the limit of detection.

The Working Group agreed that each project will have its own signal/noise optimum, and analysis methods that prove best for one project might prove unsuitable for another. Ideally, a signal/noise ratio should be optimized for each project or trial using different probe-set algorithms and data-filtering methods, and some systematic efforts towards this end are beginning to appear in the literature[17].

> "... adherence to the above QC criteria ... should yield data that are consistent between laboratories and intrinsically comparable."

After a signal is derived for each probe set, data is interpreted using statistical and visualization methods. All statistical methods run into two generic problems when faced with microarray data that are inter-related. The first is the curse of dimensionality — each gene is potentially related to every other gene, so all permutations of all available data must be considered, leading to an exponentially increasing number of possible associations in multidimensional space. The problem arises when associations (samples) become lost as the dimensionality increases — associations lose their local value and become generically global in statistical terms. Statistical models attempt to circumvent this curse by requesting larger and larger sample sizes, but fulfilling the requests becomes functionally impossible for the experimentalist. There is no easy answer to these problems and they remain a challenge for future bioinformatics research that uses microarrays[18].

Derivation of signal: probe-set algorithms and normalizations. One of the key advantages of the Affymetrix platform is the multiple measurements that are intrinsic to the probe set — most probes include 11 perfect-match and 11 paired-mismatch 25-bp oligonucleotides per gene (FIG. 1). Previous versions of GeneChip arrays used probe-set design methods that led to considerable overlap between probes, so that hybridizations to each feature/probe were not independent measurements; this led to considerable uncontrolled weighting of the contribution of any particular region of sequence to the resulting signal. All recent chips use a much more refined probe-set design with less overlap and considerably better performance of the probe set. Improvements in array and probe-set designs have been accompanied by an evolution in primary analysis algorithms and the supporting software provided by Affymetrix for data analysis and interpretation[19]. Affymetrix default algorithms are based on well-documented statistical methods, namely the robust TUKEY'S BI-WEIGHT ESTIMATOR and WILCOXON'S SIGNED RANK, to calculate the final probe-set signals and associated *p* values, respectively[19,20].

Table 1 | **Comparisons of probe-set analysis algorithms**

Algorithm	Penalty for mismatch signal	Normalization method	Outlier detection and correction	Sensitivity*	Specificity‡
Affymetrix MAS 5.0	High	Individual chips	Little	Good	Excellent
dCHIP difference model	High	Cross-project	Moderate	Good	Excellent
dCHIP	None	Cross-project	Moderate	Excellent	Good
RMA	None	Cross-project	Moderate	Excellent	Good
ProbeProfiler	Moderate	Extensive	Extensive	Good	Good

*Sensitivity is based primarily on ROC (receiver operating characteristic) curves of spike-in mRNA data based on published reports (see http://www.bioconductor.org)[21,23]. ‡Specificity measurements are based both on expectations from mismatch weights and published observations in experimental data sets[17,18].

Affymetrix has announced plans to continue to improve the software components of the GeneChip platform. The upcoming release of the GeneChip Operating System (GCOS) is expected to incorporate refinements in the user interface, data management and analysis algorithms. Software tools aside, the most significant development on the analysis front is arguably the decision by Affymetrix to release previously proprietary chip-design details, such as probe sequences, chip-design parameters and file APIs (applications programming interfaces). The goal is to encourage scientists to develop innovative analysis tools that can potentially derive more biological value from GeneChip expression data. The challenge of providing a constantly growing and evolving body of information associated with arrays has been solved in part with a web-based tool. The company's NETAFFX web site (see online links box) serves as the public portal for detailed information on chip design and has become a valuable resource for biological follow-up of GeneChip expression results. Third-party software developers can find additional support, including information on file APIs, through the Affymetrix Developers' Network (see online links box).

Encouraged in part by the openness of the platform and spawned by an increase in knowledge and experience in array data analysis, scientists are developing a number of alternative algorithms for probe-set analysis, with the goal to derive the best signal that is representative of the mRNA level for each gene. As each signal is relative to other signals in the experiment (both between arrays for the same gene and relative to all other genes on the array), the process of normalization is intimately tied to derivation of signal. The more commonly used alternative probe-set analysis algorithms include dCHIP[20], RMA[21] and ProbeProfiler (TABLE 1).

It is outside the scope of this article to discuss the nature of the different probe-set interpretation and normalization algorithms in depth, and the reader is referred elsewhere[22]. The algorithms differ in a number of important ways (TABLE 1). First, the PENALTY WEIGHT that is assigned to the mismatch probe varies — MAS 5.0 assigns a relatively heavy penalty for cross-hybridization to the mismatch probe, RMA assigns no weight and dCHIP gives the choice of providing weight or no weight. Second, the ability to discard outlier signals varies from package to package, with dCHIP and ProbeProfiler having refined methods to detect outliers at each level of analysis (probe, probe set and microarray). These packages are able to replace deviant probes with expected data based on the remainder of the probe set, and/or flag abnormal probe sets and arrays for possible exclusion from further data analysis. Third, the method of normalization varies from within a single array (MAS 5.0) to a project-based normalization (dCHIP, RMA and ProbeProfiler). Finally, MAS 5.0 provides a detection *p* value, in which a number is assigned to the confidence of the signal in question. This can be used to weight different probe-set signals in subsequent data interpretations.

> "...robust feature selection for the purpose of diagnosis and molecular markers in clinical trials requires robust statistical methods..."

The output of all packages is a normalized signal (with or without an associated detection *p* value) for each probe set on each array. These signals are then fed into data interpretation packages for statistical analyses and data visualizations.

Different probe-set interpretation algorithms lead to different results. Members of the Working Group often encounter ~50% concordance in general data output in their own work between comparisons of two different algorithms. However, it is crucial to note that the large majority of discordant data lies in regions of relatively poor signal/noise ratios, and concordance deteriorates in experiments with high levels of confounding noise. In general, the programmes that put less weight on the mismatch show better sensitivity (linearity) when signals are noisy (TABLE 1). However, this increased sensitivity can come at a cost of substantial contaminating noise[17].

The Working Group recommends using at least two probe-set algorithms for comparison and prioritization of gene selection (for example, MAS 5.0 and the dCHIP difference model).

Data interpretation. Most published microarray papers could be considered data-poor in terms of replicates and systematic statistical analyses, but data-rich both in terms of amount of high-quality data generated and significant research findings. Below, we point out the most appropriate current bioinformatics methods and additional methods that require further development so that data can be more fully mined for information content.

A second general backdrop to the following discussion is that data visualization is one of the most powerful data interpretation tools, yet it rarely obeys statistical principles. The resolution of the human eye, coupled with the abstract computational power of the human brain, lies behind the popularity of hierarchical clustering and other non-statistical principles and visualization methods. However, the eye and brain are poorly suited to spontaneously deriving statistical support.

There are two general types of experimental design that lend themselves to different types of statistical and visual analysis: the cross-sectional study and the TIME-SERIES STUDY. The cross-sectional study typically has gene or pattern selection as the goal: the identification of one or more genes or patterns of expression that are diagnostic of the condition or state

under study. This 'gene selection' might be for truly diagnostic purposes (for example, differential diagnosis of leukaemia), or might be intended to identify relevant biochemical pathways. In both cases, the gene or pattern selection must be robust, usually implying a statistically principled approach, with subsequent validation by predictive computer modelling (internal cross-validation) or, preferably, prospective validation on new data.

Feature selection can be the main limiting factor in evaluation of the predictive performance of an analysis method when there are many predictors to select from. This was a 'mantra' for some of the senior statisticians involved in predictive modelling with gene-expression array data for several years, but only now do the non-statistical users and developers of predictive models from non-statistical perspectives begin to appreciate these issues. Proper validation of any model or algorithm that relies on explicit feature selection — such as choosing a subset of 70 genes from 20,000 — that underlies the resulting prediction simply must ensure that the analysis is tested by internal cross-validation that includes feature re-selection as part of the validation[23,24]. The Working Group acknowledged that prospective validation of any findings using new data is the acid test of predictive performance. The focus on feature or gene selection is vitally important when microarrays are used for differential diagnosis and has been best studied in cancer biopsy/tissue studies.

An increasing proportion of microarray studies focus on delineation of biochemical pathways that are modulated in response to some stimulus. In practice, these studies typically use feature selection to identify potential pathways that are involved in the response of the cells or tissues. Validation is then done on the identified biochemical pathways of interest, using mRNA (real-time PCR) or protein studies, often proving cause and effect in experimental models.

The Working Group notes that robust feature selection for the purpose of diagnosis and molecular markers in clinical trials requires robust statistical methods, as outlined below, and the burden of proof lies with statistical validation. For microarray experiments designed to delineate biochemical pathways, feature selection is used for generating a hypothesis and the burden of proof of the hypothesis lies with laboratory-based research, often at the protein level.

For feature selection, the Working Group recommends that users experiment with various statistical methods (such as standard parametric tests, nonparametric methods, false discovery rate and related methods[25], global or local shrinkage of raw signal intensities and Stanford's 'nearest shrunken centroids'[26]). Developments related to SURVIVAL DATA ANALYSIS are receiving increased attention because clinical trials will raise the need to move that way. As a corollary, analysis methods that focus on signatures of groups of genes (such as averages of clusters, Duke's metagenes[27–29] and Stanford's eigengenes[30]) seem worth stressing in predictive contexts.

Glossary

A-, B- AND C-SERIES ARRAYS
A series of human, rat and mouse Affymetrix arrays released in 2003, in which the A array contained the best-characterized genes, and B and C arrays contained less well-defined expressed sequence tags. In 2004, all probe sets have been condensed so that there is only one microarray per species that covers the entire genome.

CROSS-SECTIONAL DESIGN
The use of different subjects in an experimental and control group or groups. The statistical analysis compares the median and variation within each group relative to the other groups.

FEATURE
Typically one element (spot) on a microarray. In spotted cDNA or oligonucleotide arrays, features correspond to genes or transcripts; in Affymetrix arrays, there are typically 22 elements per probe set and often multiple probe sets per gene, so a feature might refer to a single oligonucleotide, a probe pair or a probe set, or a gene with multiple probe sets. In bioinformatics it is most often synonymous with a gene.

FLUORESCENCE-ACTIVATED CELL SORTING (FACS)
A method whereby dissociated and individual living cells are sorted, in a liquid stream, according to the intensity of fluorescence that they emit as they pass through a laser beam.

FLUOROPHORE
A small molecule, or a part of a larger molecule, that can be excited by light to emit fluorescence.

ISCHAEMIA
The loss of blood supply, and hence oxygenation, to a tissue or cells.

LASER CAPTURE MICRODISSECTION
A technique in which individual cells, or regions of tissue, are excised from a histological preparation, using specially equipped microscopes, and isolated for further study.

LONGITUDINAL DESIGN
The use of multiple samples from the same subject. With this design, each subject serves as their own control, eliminating confounding inter-individual variations at baseline; paired *t*-tests are used to interpret the data.

NEGATIVE CELL ISOLATION
The use of antibodies or other reagents to remove all unwanted cells from a mixed population of cells. In this method, the desired cells are not exposed to bound antibodies, thereby avoiding potential activation or other molecular alteration in the desired cells.

PENALTY WEIGHT
In Affymetrix arrays, hybridization to the 'mismatch' probe of a probe pair might or might not be considered as a form of measurement of noise or background, and can be factored into the signal seen with the paired 'perfect match' as a penalty weight.

PHOTOLITHOGRAPHY
The process of using light to either etch or activate regions of a surface (substrate). This method is used in microelectronics to create integrated circuits and processors.

REAL-TIME PCR
The quantification of the amount of PCR product during each cycle of a PCR reaction. The product concentration, as a function of cycle number, provides a good estimation of the relative quantity of the mRNA being tested.

RESECTION
Surgical removal of tissue, most commonly used for removing tumorous masses from surrounding tissue.

S1 NUCLEASE PROTECTION
An experimental method for determining mRNA transcript concentration in a tissue or cell RNA sample. It involves using labelled DNA probes that bind the RNA, with overhanging non-hybridized tails of the probe then being digested by the S1 nuclease. This creates a smaller labelled DNA probe that is indicative of the abundance of the mRNA being tested.

SURVIVAL DATA ANALYSIS
A battery of statistical methods applied to data when mortality is often the only, or best, measured outcome.

TIME-SERIES STUDY
The use of a series of samples taken at defined time points after a defined stimulus. In mice and rats, the samples at different time points are usually from different animals. In humans, time-series studies are necessarily longitudinal to avoid additional confounding noise.

TUKEY'S BI-WEIGHT ESTIMATOR
Many statistical tests require underlying definitions that are assumed to be valid (for example, tumour versus non-tumour), and require data that show a normal distribution. Microarray data, and the clinical information underlying the definition of samples, is often less exact, with genes or samples often performing as statistical outliers. Tukey's bi-weight estimator is one of the M-class of statistical models that is less sensitive to outliers and performs more gracefully when underlying assumptions are inexact.

WILCOXON'S SIGNED RANK
A statistical test that investigates the population median of paired differences. It is well suited for microarray work as it treats each gene as an independent variable and does not require normal distributions of the data.

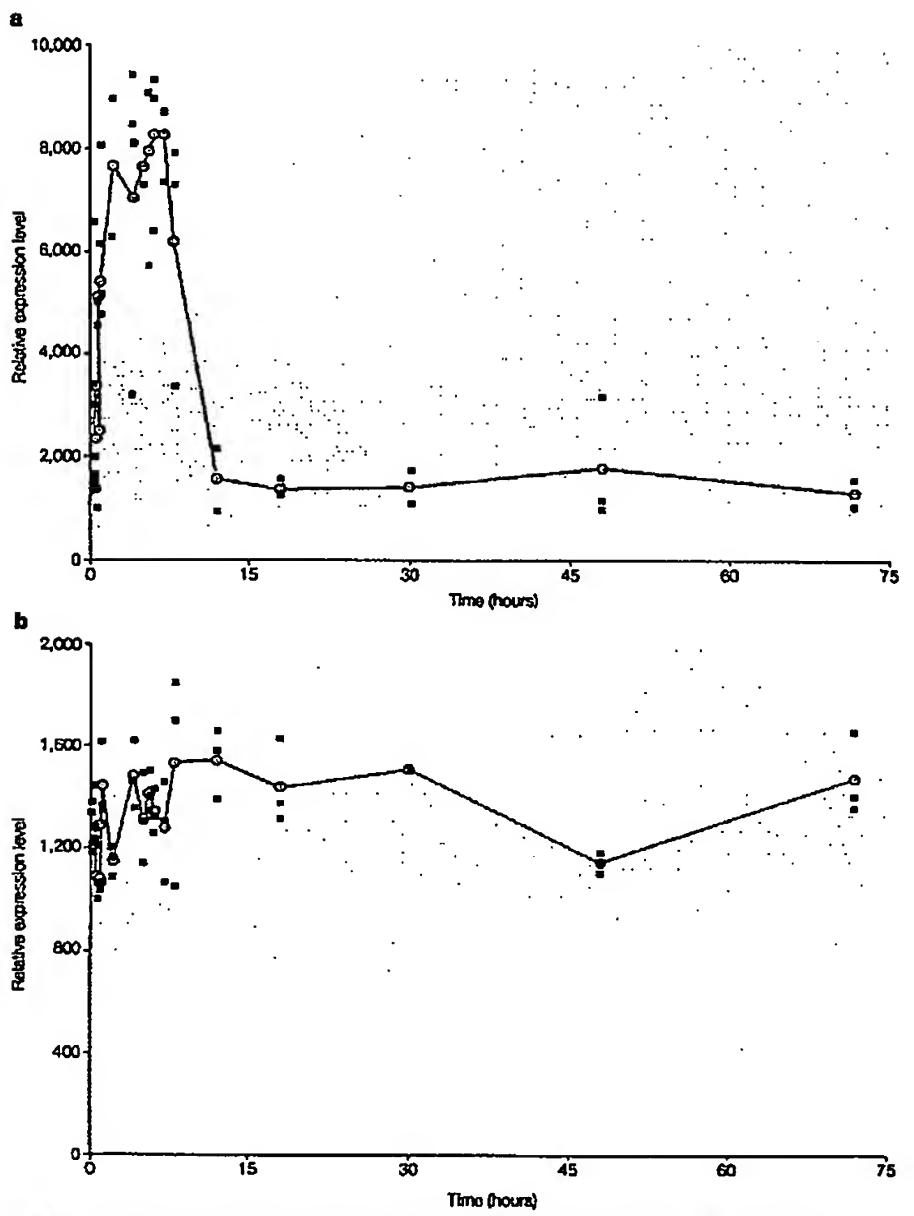


Figure 2 | Dense time-series data with adequate replicates can provide robust visual interpretation of data. Dynamic, single-gene queries can provide visually compelling results and avoid many issues that complicate statistical analyses of cross-sectional microarray studies. Dynamic queries of the *got1* transcript probe set from a time-series study is shown. The *x*-axis represents time (in hours) after rats were given a bolus of 3-methylprednisolone. The *y*-axis represents relative expression level. Individual rats were studied at each time point (green data points), with both liver (panel a) and muscle (panel b) tissue taken from the same animals. Averages of the replicates for each data point are shown (in magenta); the graph line is drawn between the averages. At baseline (time 0), the *got1* transcript has a normalized signal of about 2,400 in liver and 1,200 in muscle. The gene is clearly responsive to 3-methylprednisolone in liver (panel a), where expression rapidly increases within the first hour, plateaus between 1–7 hr, then quickly falls back to baseline by 12 hr. The replicates appear relatively consistent. By contrast, the same gene in muscle does not seem to respond to the drug; the variability in replicates is larger than any temporally-relevant change. Data from http://microarray.cnmcresearch.org/singlegenemain.asp and REFS 6,32.

Whatever the specific statistical model that is applied for prediction, using aggregate gene expression has important consequences: measures of aggregation of expression over a group of genes with related profiles can reduce dimension (thereby mitigating the curse of dimensionality). This can reduce multiplicities and, to some degree, ease the problems of gene selection, multiple testing and co-linearity, while improving signal estimation by averaging correlated noise components.

Data visualizations, time series and candidate genes. The above discussions of biostatistics all assume that the analysis is targeted towards a cross-sectional study, in which the primary goal is diagnostic gene discovery (gene or feature selection). In other words, a series of microarrays with a very large number of transcripts defines the very small minority of genes that are correlated and therefore predictive of the biological variable of interest. There are alternatives to this standard experimental design that use entirely different types of analysis, and the statistical issues are also quite different, as explained below.

The time-series study, if done with enough time points, can provide an effective antidote to the curse of dimensionality — the action of any gene during a time-series study should make biological sense, such that each signal is relatively easily discernible from noise. Visual query of a large time-series data set for single gene responses to the controlled variable either might meet expectations and is therefore valid, or might not meet expectations and is discarded as uninteresting. As an example, we show a time-series study in which rats are given a bolus of methylprednisolone, after which their liver and muscle are studied as a function of time (FIG. 2). In this case, the same gene (*got1*) is queried using a web-based dynamic visualization tool, first in liver (FIG. 2a) and then in muscle (FIG. 2b). The data in the top panel are visually compelling; *got1* in liver responds quickly and strongly to a bolus of 3-methylprednisolone, with relatively consistent replicates (each data point comes from a different animal) and a time course that is visually assuring so that complex statistical tests of the transcriptional response as a function of time are not needed. On the other hand, the same gene in muscle does not seem to respond to the drug[6,31] (FIG. 2b). Through such gene queries, the variability in replicates and the appropriateness of the action of the gene as a function of time can quickly be assessed. Another advantage of time-series data is that such profiles act as biomarkers that are amenable to analysis and interpretation using pharmacodynamic models that predict the underlying mechanisms of control of gene expression[32].

The Working Group agreed that data-rich, time-series experimental designs provide some latitude in reporting significant findings and that the query of individual genes within large data sets can circumvent complex issues of multidimensionality of data.

Future areas of development

The data-rich and highly dimensional nature of microarray data serves as a model for future dissection and understanding of biological

systems in general, including proteomics and integration of mRNA profiling and proteomics. The Working Group discussed data analysis needs within the microarray community and agreed that, along with the incorporation of QC, SOPs and optimized or customized signal/noise analyses in initial project signal generation, the back-end statistics needs to reach a commonly accepted method of dealing with the curse of dimensionality before microarrays can be reliably used in clinical trials. Statisticians need to focus more on representation of prediction results in terms of probabilities and associated measures of uncertainty, and reach a consensus on what is acceptable. In the meantime, it is likely that specific marker or diagnostic genes will be extracted from pilot profiling studies, and then only this small subset of genes will be used as a clinical trial endpoint. This data limitation approach removes much of the curse of dimensionality, but is liable to ignore the large majority of data, thereby decreasing the potential power of the study and bringing into question the use of microarrays in clinical trials.

A move towards the standardization of reporting of prediction accuracy would be desirable when assessing predictive accuracy through within-sample cross-validation. The Working Group suggests that one or more validation techniques be used when reporting predictive genes: leave-one-out and 10% cross-validation summaries, or true validation data sets. Communicating uncertainty about predictive performance is also key and will help evaluate results based on varying sample sizes. The Working Group suggests that until this information is routinely presented in published papers, it will be difficult to reach an acceptable consensus for use in clinical trials.

Conclusions

There are four key areas of optimization and standardization that are largely independent: study design, technical variability (QC/SOP of data generation), analysis method variation (signal/noise optimization using probeset algorithms and normalizations) and back-end statistical analyses. Statistics of clinical trial design is crucial: gene-expression data does not mitigate the need for sound and relevant design and analysis, nor does it challenge what we know about design. The field is quickly maturing from the small-chip-number hit-and-run type projects to those with a more robust study design. However, study design depends ultimately on appropriate powering of a study, which is greatly affected by both the chip-analysis algorithms that are used and the biostatistical data analysis.

Development of back-end statistical methods for data representation/summary and for high-level analysis remains an active area of research for both academic and commercial users, and is likely to remain so in the near future. We are some way from defining standards of summary signal intensities alone and even further from considerations of standardization of analytical methods for inference and prediction in clinical contexts. In regulated clinical studies, such standards will be enforced partly by the US FDA as submissions of medical test/device protocols emerge and increase in number. Even then, however, many approaches to data analysis and modelling will be used and developed, which is, of course, to be supported. It is very difficult to influence the research community, especially when the variety of problems that are encountered promotes the need for refined and new approaches.

*Members of The Tumor Analysis Best Practices Working Group are listed in Box 3.
Correspondence to Eric P. Hoffman at the Center for Genetic Medicine, Children's National Medical Center, 111 Michigan Avenue NW, Washington DC 20010, USA.
e-mail: ehoffman@cnmcresearch.org

doi:10.1038/nrg1297

1. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genet.* 29, 365–371 (2001).
2. Spellman, P. T. *et al.* Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol.* 3, RESEARCH0046.1–0046.9 (2002).
3. Brazma, A. *et al.* ArrayExpress — a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 31, 68–71 (2003).
4. Zhao, P., Iezzi, S., Sartorelli, V., Dressman, D. & Hoffman, E. P. Slug is downstream of myoD: identification of novel pathway members via temporal expression profiling. *J. Biol. Chem.* 277, 20091–20101 (2002).
5. Di Giovanni, S. *et al.* Gene profiling in spinal cord injury shows role of cell cycle in neuronal death. *Ann. Neurol.* 53, 454–468 (2003).
6. Jin, J. Y., Almon, R. R., DuBois, D. C. & Jusko, W. J. Modeling of corticosteroid pharmacogenomics in rat liver using gene microarrays. *J. Pharmacol. Exp. Ther.* 307, 93–109 (2003).
7. Bakay, M. *et al.* Sources of variability and effect of experimental approach on expression profiling data interpretation. *BMC Bioinformat.* 3, 4–15 (2002).
8. DePrimo, S. E. *et al.* Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. *BMC Cancer* 3, 3 (2003).
9. de Vos, S. *et al.* Gene expression profile of serial samples of transformed B-cell lymphomas. *Lab. Invest.* 83, 271–285 (2003).
10. Hittel, D. S., Kraus, W. E. & Hoffman, E. P. Skeletal muscle dictates the fibrinolytic state after exercise training in overweight men with characteristics of metabolic syndrome. *J. Physiol.* 548, 401–410 (2003).
11. Zambon, A. C. *et al.* Time- and exercise-dependent gene regulation in human skeletal muscle. *Genome Biol.* 4, R61 (2003).
12. Bakay, M. *et al.* Sources of variability and effect of experimental approach on expression profiling data interpretation. *BMC Bioinformat.* 3, 4–15 (2002).
13. Cardozo, A. K. *et al.* Gene microarray study corroborates proteomic findings in rodent islet cells. *J. Proteome Res.* 2, 553–555 (2003).
14. Chun, T. W. *et al.* Gene expression and viral production in latently infected, resting CD4+ T cells in viremic versus aviremic HIV-infected individuals. *Proc. Natl Acad. Sci. USA* 100, 1908–1913 (2003).
15. Komma, F. *et al.* Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J. Neurosci.* 23, 3607–3615 (2003).
16. Huang, J. *et al.* Effects of ischemia on gene expression. *J. Surg. Res.* 99, 222–227 (2001).
17. Seo, J. *et al.* Interactive color mosaic and dendrogram displays for signal/noise optimization in microarray data analysis. *IEEE ICME* 3, 461–462 (2003).
18. Somorjai, R. L., Dolenko, B. & Baumgartner, R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics* 19, 1484–1491 (2003).
19. Mei, R. *et al.* Probe selection for high-density oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* 100, 11237–11242 (2003).
20. Li, C. & Hung Wong, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol.* 2, RESEARCH0032 (2001).
21. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31, e15 (2003).
22. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193 (2003).
23. Ambroise, C. & McLachlan, G. J. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc. Natl Acad. Sci. USA* 99, 6562–6566 (2002).
24. West, M. *et al.* Predicting the clinical status of human breast cancer utilizing gene expression profiles. *Proc. Natl Acad. Sci. USA* 98, 11462–11467 (2001).
25. Tusher, V., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* 98, 5116–5124 (2001).
26. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl Acad. Sci. USA* 99, 6567–6572 (2002).
27. Huang, E. *et al.* Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nature Genet.* 34, 226–230 (2003).
28. Black, E. P. *et al.* Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res.* 63, 3716–3723 (2003).
29. Huang, E. *et al.* Gene expression predictors of breast cancer outcomes. *Lancet* 361, 1590–1596 (2003).
30. Alter, O., Brown, P. O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci. USA* 97, 10101–10106 (2000).
31. Chen, J. *et al.* The PEPR GeneChip data warehouse and implementation of a dynamic time series query tool (SGQT) with graphical interface. *Nucleic Acids Res.* 32, D578–D581 (2004).
32. Almon *et al.* In vivo multitissue corticosteroid microarray time series available online at Public Expression Profile Resource (PEPR). *Pharmacogenomics* 4, 791–799 (2003).

Acknowledgements

The authors thank their respective funding agencies, particularly the larger collaborative funding initiatives that make systematic and large-scale studies of the bioinformatics and biostatistics of genome-wide data sets possible from the Department of Defense and the Doris Duke Charitable Foundation CSDA. The authors also thank S. Hilmer, A. DeBiase and G. Miyada for their critique of the manuscript.

Competing interests statement

Some of the authors declare competing financial interests: see Web version for details.

Online links

DATABASES

The following terms in this article are linked online to:
LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink
GAPDH | *got1*

FURTHER INFORMATION

Affymetrix Developers' Network:
http://www.affymetrix.com/support/developer/index.affx
Affymetrix Technical Note:
http://www.affymetrix.com/support/technical/technotes/blood_technote.pdf
Agilent Technologies: http://www.chem.agilent.com
ArrayExpress microarray database:
http://www.ebi.ac.uk/arrayexpress
The Children's National Medical Center Microarray Center:
http://microarray.cnmcresearch.org/singlegenemain.asp
HOPGENE Program for Genomic Applications:
www.hopkins-genomics.org
MGED Data Transformation and Normalization Working Group: http://www.dnachip.org/mged/normalization.html
MGED Society: http://www.mged.org
NETAFFX web site:
http://www.affymetrix.com/analysis/index.affx
NIGMS Glue Grant:
http://www.gluegrant.org/whatsagluegrant.htm
Access to this interactive links box is free online.

EXHIBIT D

MOLECULAR BIOLOGY OF THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

GS Garland Science
Taylor & Francis Group

Garland
Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilernia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

Library of Congress Cataloging-in-Publicaton Data
Molecular biology of the cell / Bruce Alberts ... [et al.].-- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group, 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Cell Biology Interactive
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Front cover Human Genome: Reprinted by permission from *Nature,* International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila, Arabidopsis,* Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

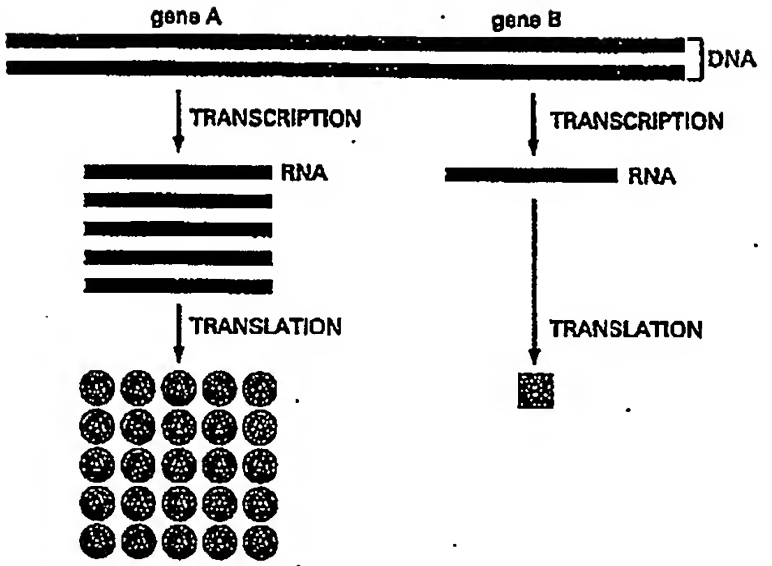


Figure 6–3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6–3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription**.

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6–4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribo*nucleic acid) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6–5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6–6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

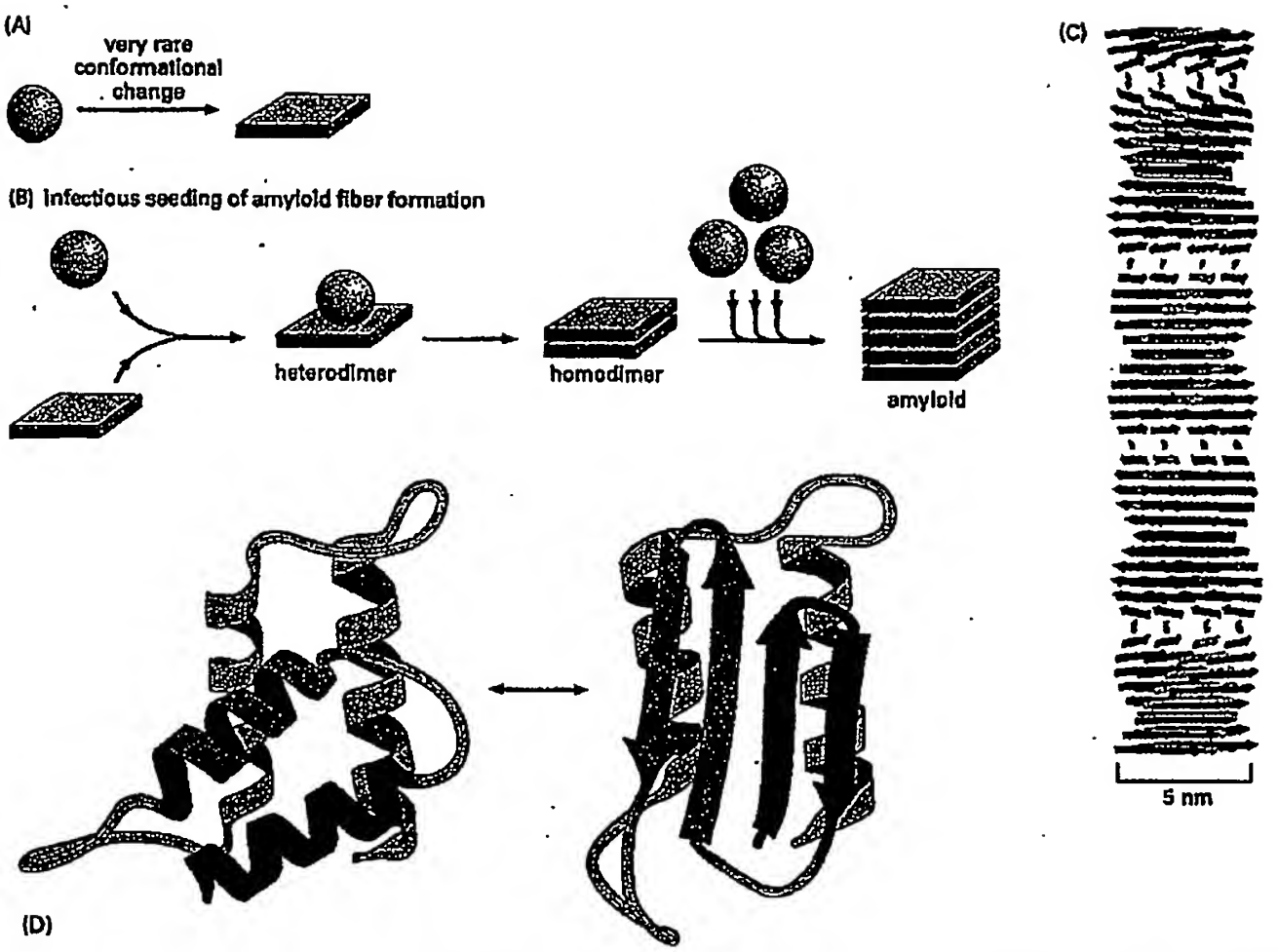


Figure 6–89 Protein aggregates that cause human disease. (A) Schematic illustration of the type of conformational change in a protein that produces material for a cross-beta filament. (B) Diagram illustrating the self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP*, induces the normal PrP protein it contacts to change its conformation, as shown. Most of the human diseases caused by protein aggregation are caused by the overproduction of a variant protein that is especially prone to aggregation, but because this structure is not infectious in this way, it cannot spread from one animal to another. (C) Drawing of a cross-beta filament, a common type of protease-resistant protein aggregate found in a variety of human neurological diseases. Because the hydrogen-bond interactions in a β sheet form between polypeptide backbone atoms (see Figure 3–9), a number of different abnormally folded proteins can produce this structure. (D) One of several possible models for the conversion of PrP to PrP*, showing the likely change of two α-helices into four β-strands. Although the structure of the normal protein has been determined accurately, the structure of the infectious form is not yet known with certainty because the aggregation has prevented the use of standard structural techniques. (C, courtesy of Louise Serpell, adapted from M. Sunde et al., *J. Mol. Biol.* 273:729–739, 1997; D, adapted from S.B. Prusiner, *Trends Biochem. Sci.* 21:482–487, 1996.)

animals and humans. It can be dangerous to eat the tissues of animals that contain PrP*, as witnessed most recently by the spread of BSE (commonly referred to as the "mad cow disease") from cattle to humans in Great Britain.

Fortunately, in the absence of PrP*, PrP is extraordinarily difficult to convert to its abnormal form. Although very few proteins have the potential to misfold into an infectious conformation, a similar transformation has been discovered to be the cause of an otherwise mysterious "protein-only inheritance" observed in yeast cells.

There Are Many Steps From DNA to Protein

We have seen so far in this chapter that many different types of chemical reactions are required to produce a properly folded protein from the information contained in a gene (Figure 6–90). The final level of a properly folded protein in a cell therefore depends upon the efficiency with which each of the many steps is performed.

We discuss in Chapter 7 that cells have the ability to change the levels of their proteins according to their needs. In principle, any or all of the steps in Fig-

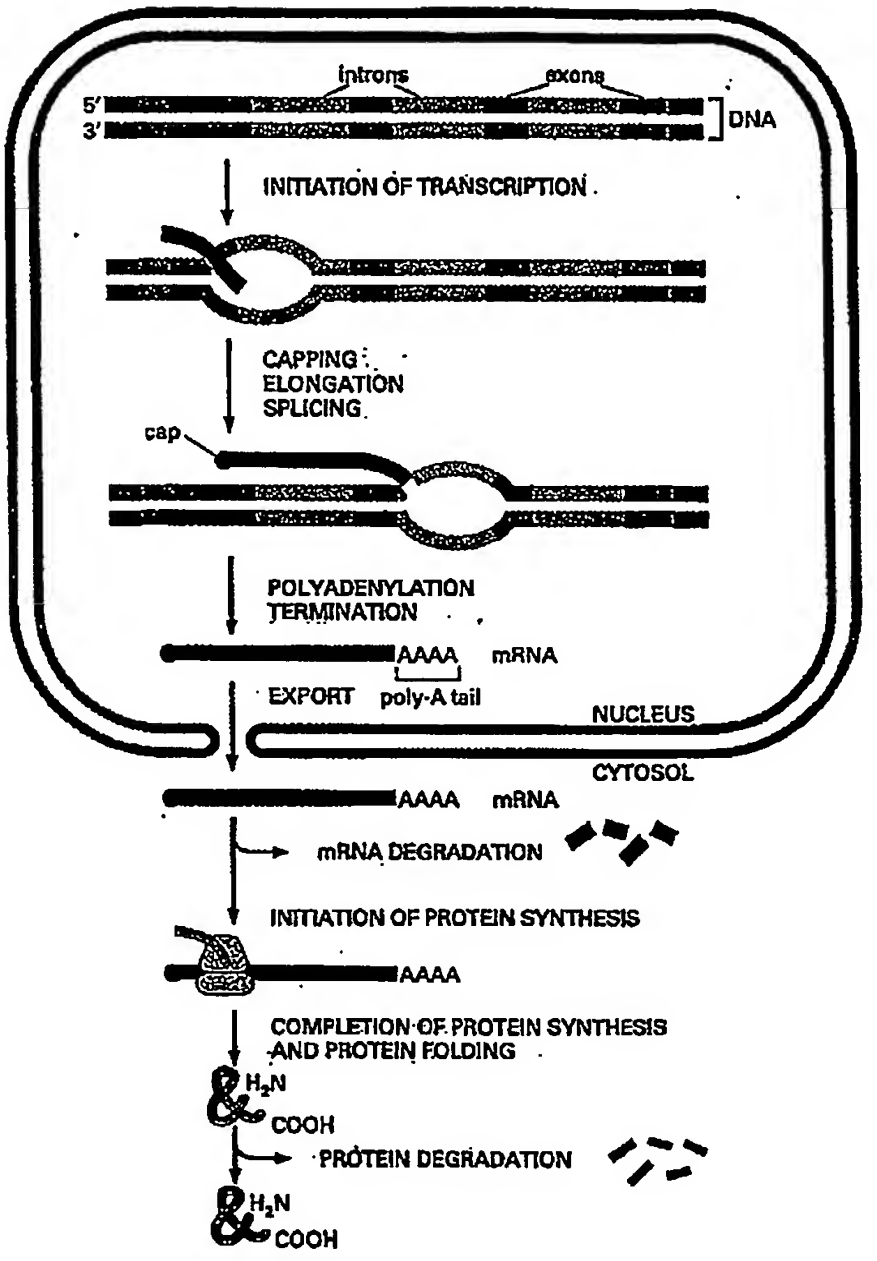


Figure 6–90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6–90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

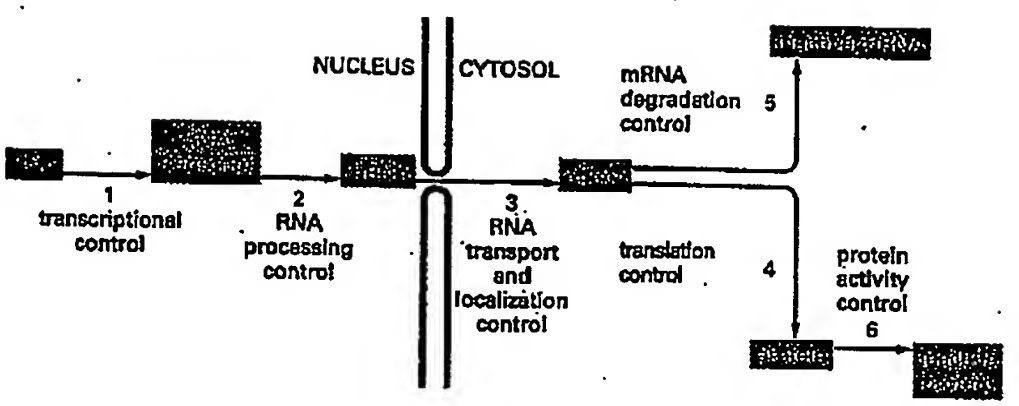


Figure 7–5 Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7–5).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7–5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices

occur in the germ line, the cell lineage that gives rise to sperm or eggs. Most of the DNA in vertebrate germ cells is inactive and highly methylated. Over long periods of evolutionary time, the methylated CG sequences in these inactive regions have presumably been lost through spontaneous deamination events that were not properly repaired. However promoters of genes that remain active in the germ cell lineages (including most housekeeping genes) are kept unmethylated, and therefore spontaneous deaminations of Cs that occur within them can be accurately repaired. Such regions are preserved in modern day vertebrate cells as CG islands. In addition, any mutation of a CG sequence in the genome that destroyed the function or regulation of a gene in the adult would be selected against, and some CG islands are simply the result of a higher than normal density of critical CG sequences.

The mammalian genome contains an estimated 20,000 CG islands. Most of the islands mark the 5′ ends of transcription units and thus, presumably, of genes. The presence of CG islands often provides a convenient way of identifying genes in the DNA sequences of vertebrate genomes.

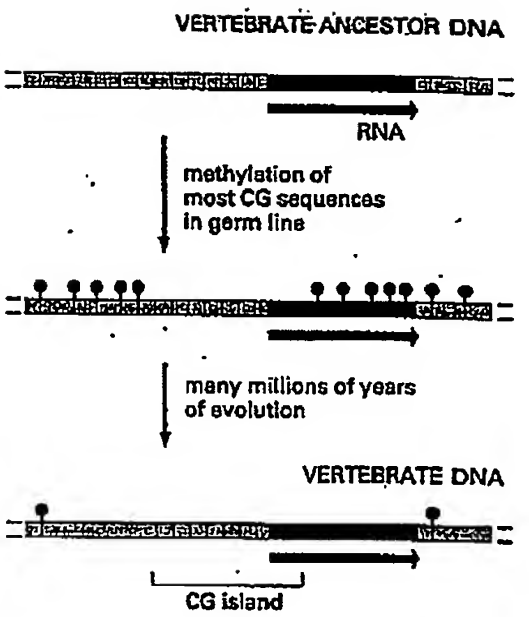


Figure 7–86 A mechanism to explain both the marked overall deficiency of CG sequences and their clustering into CG islands in vertebrate genomes. A *black line* marks the location of a CG dinucleotide in the DNA sequence, while a red "lollipop" indicates the presence of a methyl group on the CG dinucleotide. CG sequences that lie in regulatory sequences of genes that are transcribed in germ cells are unmethylated and therefore tend to be retained in evolution. Methylated CG sequences, on the other hand, tend to be lost through deamination of 5-methyl C to T, unless the CG sequence is critical for survival.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character through many cell division cycles and even when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides. These features endow the cell with a memory of its developmental history. Bacteria and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms. One such mechanism involves a competitive interaction between two gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory. Negative feedback loops with programmed delays form the basis for cellular clocks.

In eucaryotes the transcription of a gene is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be active in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also used by eucaryotic cells to regulate gene expression. An especially dramatic case is the inactivation of an entire X chromosome in female mammals. In vertebrates DNA methylation also functions in gene regulation, being used mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms. DNA methylation also underlies the phenomenon of genomic imprinting in mammals, in which the expression of a gene depends on whether it was inherited from the mother or the father.

POST-TRANSCRIPTIONAL CONTROLS

In principle, every step required for the process of gene expression could be controlled. Indeed, one can find examples of each type of regulation, although any one gene is likely to use only a few of them. Controls on the initiation of gene transcription are the predominant form of regulation for most genes. But other controls can act later in the pathway from DNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls**, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control*, for many genes they are crucial.

Exhibit 2

GNE.3230R1C39 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Eaton, et al.
Appl. No.	:	10/063,557
Filed	:	May 2, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	David J. Blanchard
Group Art Unit	:	1642

DECLARATION OF J. CHRISTOPHER GRIMALDI, UNDER 37 C.F.R. § 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, J. Christopher Grimaldi, declare and say as follows:

1. I am a Senior Research Associate in the Molecular Biology Department of Genentech, Inc., South San Francisco, CA 94080.

2. I joined Genentech in January of 1999. From 1999 to 2003, I directed the Cloning Laboratory in the Molecular Biology Department. During this time I directed or performed numerous molecular biology techniques including qualitative Polymerase Chain Reaction (PCR) analyses. I am currently involved in, among other projects, the isolation of genes coding for membrane associated proteins which can be used as targets for antibody therapeutics against cancer. In connection with the above-identified patent application, I personally performed or directed the semi-quantitative PCR analyses in the assay entitled "Tumor Versus Normal Differential Tissue Expression Distribution" which is described in EXAMPLE 18 in the specification that were used to identify differences in gene expression between tumor tissue and their normal counterparts.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

4. In differential gene expression studies, one looks for genes whose expression levels differ significantly under different conditions, for example, in normal versus diseased tissue.

Chromosomal aberrations, such as gene amplification, and chromosomal translocations are important markers of specific types of cancer and lead to the aberrant expression of specific genes and their encoded polypeptides, including over-expression and under-expression. For example, gene amplification is a process in which specific regions of a chromosome are duplicated, thus creating multiple copies of certain genes that normally exist as a single copy. Gene under-expression can occur when a gene is not transcribed into mRNA. In addition, chromosomal translocations occur when two different chromosomes break and are rejoined to each other chromosome resulting in a chimeric chromosome which displays a different expression pattern relative to the parent chromosomes. Amplification of certain genes such as Her2/Neu [Singleton *et al.*, Pathol. Annu., 27Pt1:165-190], or chromosomal translocations such as t(5;14), [Grimaldi *et al.*, Blood, 73(8):2081-2085(1989); Meeker *et al.*, Blood, 76(2):285-289(1990)] give cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy. When the chromosomal aberration results in the aberrant expression of a mRNA and the corresponding gene product (the polypeptide), as it does in the aforementioned cases, the gene product is a promising target for cancer therapy, for example, by the therapeutic antibody approach.

5. Comparison of gene expression levels in normal versus diseased tissue has important implications both diagnostically and therapeutically. For example, those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed, as evidenced by an increased production of mRNA, the gene product or polypeptide will also be over-expressed. It is unlikely that one identifies increased mRNA expression without associated increased protein expression. This same principle applies to gene under-expression. When a gene is under-expressed, the gene product is also likely to be under-expressed. Stated in another way, two cell samples which have differing mRNA concentrations for a specific gene are expected to have correspondingly different concentration of protein for that gene. Techniques used to detect mRNA, such as Northern Blotting, Differential Display, *in situ* hybridization, quantitative PCR, Taqman, and more recently Microarray technology all rely on the dogma that a change in mRNA will represent a similar change in protein. If this dogma did not hold true then these techniques would have little value and not be so widely used. The use of mRNA quantitation techniques have identified a seemingly endless number of genes which are differentially expressed in various tissues and these genes have subsequently been shown to have correspondingly similar changes in their protein levels. Thus, the detection of increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression. The detection of increased or decreased polypeptide expression can be used for cancer diagnosis and treatment.

6. However, even in the rare case where the protein expression does not correlate with the mRNA expression, this still provides significant information useful for cancer diagnosis and treatment. For example, if over- or under-expression of a gene product does not correlate with over- or under-expression of mRNA in certain tumor types but does so in others, then identification of both gene expression and protein expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over- or

under-expression of the gene product in the presence of a particular over- or under-expression of mRNA is crucial information for the practicing clinician. For example, if a gene is over-expressed but the corresponding gene product is not significantly over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: 8/10/2004

J. Christopher Grimaldi

S:\DOCS\AOK\AOK-5479.DOC
071904

8 8

GNE2930R1C3

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Botstein, et al.
Appl. No.	:	10/032,996
Filed	:	December 27, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Fredman, J.
Group Art Unit	:	1634

DECLARATION OF PAUL POLAKIS, PH.D. UNDER 37 C.F.R. § 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Attached is the Declaration of Paul Polakis, Ph.D.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: June 14, 2004

By: AnneMarie Kaiser
AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

S:\DOCS\BSG\BSG-1215.DOC
061404

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).

2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.

4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.

5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

SV 2031808 v1

EXHIBIT A

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South SanFrancisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985	Assistant Professor, Department of Chemistry, Oberlin College, Oberlin, Ohio
1980-1984	Graduate Research Assistant, Department of Biochemistry, Michigan State University East Lansing, Michigan

PUBLICATIONS:

1. Polakis, P G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.

2. Polakis, P.G: and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys**. 234, 341-352.

3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys**. 236, 328-337.

4. Uhing, R.J., **Polakis,P.G.** and Snyderman, R. 1987 Isolaton of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.

5. Polakis, P.G., Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem**. 263, 4969-4979.

6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. Polakis, P.G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun**. 161, 276-283.

8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Comun.** 160, 25-32.

9. Polakis, P., Weber,R.F., Nevins,B., Didsbury, J. Evans,T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem**. 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.G.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Charaterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K^+ Channels. **Cell.** 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.

14. Polakis, P.G. Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits $p21^{ras}$ Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein $p21^{rap1}$. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for $p21^{ras}$-dependent Inhibition of Muscarinic Atrial K^+ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P** . 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 $p60^{v\text{-}src}$ Can Associate with and Phosphorylate the $p21^{ras}$ GTPase Activating Protein. **Oncogene** 7,389-394.

22. Polakis P. and McCormick, F. 1992 Interactions Between $p21^{ras}$ Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., **Polakis,P.** and McCormick, F. 1992 Molecular coloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.

24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle $p34^{cdc2}$ Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., **Polakis, P.**, McCormick, F., and Innis, M. A. 1992 Functional Interactions Between $p21^{rap1A}$ and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy,L., Clark, R., McCormick, F. and **Polakis, P.** 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell . Biol.** 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., **Polakis, P.**, McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.

28. Janoueix-Lerosey, I., **Polakis, P.**, Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.

29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the *ras*-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of $p21^{ras}$ with GAP, exchange factors, and its bological effector target. **J. Biol Chem.** 268, 9157-9160.

31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and **Polakis, P.** 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., **Polakis, P.**, McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.

33. Sato, K. Y., **Polakis, P.**, Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor acitvity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., **Polakis, P.** and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and **Polakis, P.** 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31

37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71

38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α-catenin, β-catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555

39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β-catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.

40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.

41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.

42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.

43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β-catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.

44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein witha rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.

45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.

46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.

47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635

48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in Xenopus embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β-catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P** 1997 Loss of β-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes. K., Waterman, M., and **Polakis, P.** 1997 Induction of a β-catenin-LEF-1 complex by wnt-1 and transforming mutants of β-catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ.** 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β-catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β-catenin and glycogen synthase kinase 3β. **Current Biology 8,** 573-581.

56. Polakis, P. 1998 The oncogenic activation of β-catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfled, Florence Margottin, Richard Benarous and **Paul Polakis**. 1999 The F-box protein β–TrCP associates with phosphorylated β–catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β–catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P,** Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β-catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.

61 **Polakis P**, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32

62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.

65. Peifer M, **Polakis P.** 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9,

66. **Polakis P.** 2000 Wnt signaling and cancer. **Genes Dev;**14, 1837-1851.

67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.

68. Szeto , W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.

69. Rubinfeld B, Tice DA, **Polakis P.** 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.

70. **Polakis P.** 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.

71. Tice DA, Soloviev I, **Polakis P.** 2002 Activation of the Wnt Pathway Interferes withSerum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem.** 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennnica D, **Polakis P.** 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.

73. **Polakis, P.** 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao,W. , Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., **Polakis, P.** 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, **Polakis P,** Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics,** April 8

MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts . . . [et al.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]
QH581.2.M64 1994
574.87—dc20
DNLM/DLC
for Library of Congress 93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 10 9 8 7

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)


carrot

a whole
lated

ability to
igle cell
cells that
e plant.

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

A Cell Can Change the Expression of Its Genes in Response to External Signals [3]

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein [4]

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9–2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9–2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

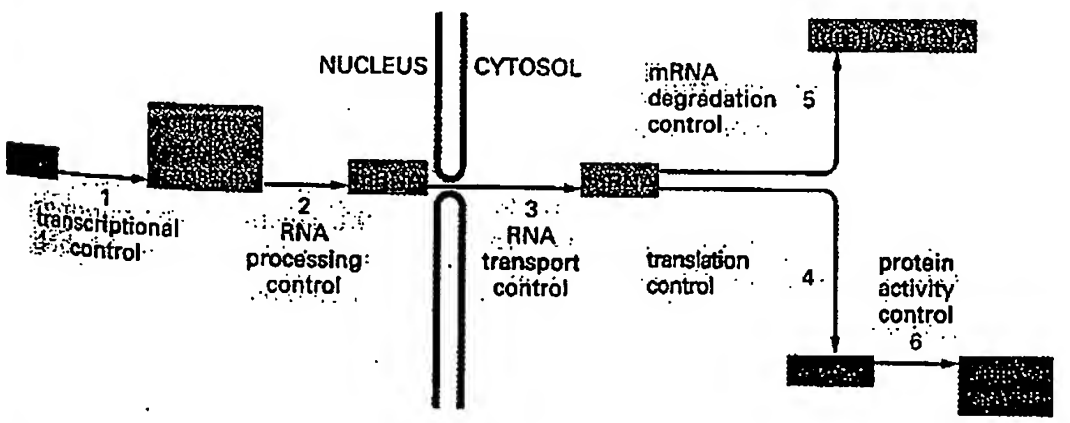


Figure 9–2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

following sections we discuss the DNA and protein components that regulate the initiation of gene transcription. We return at the end of the chapter to the other ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-binding Motifs in Gene Regulatory Proteins [5]

How does a cell determine which of its thousands of genes to transcribe? As discussed in Chapter 8, the transcription of each gene is controlled by a regulatory region of DNA near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Other regulatory regions are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices consist of two fundamental types of components: (1) short stretches of DNA of defined sequence and (2) *gene regulatory proteins* that recognize and bind to them.

We begin our discussion of gene regulatory proteins by describing how these proteins were discovered.

Gene Regulatory Proteins Were Discovered Using Bacterial Genetics [6]

Genetic analyses in bacteria carried out in the 1950s provided the first evidence of the existence of **gene regulatory proteins** that turn specific sets of genes on or off. One of these regulators, the *lambda repressor,* is encoded by a bacterial virus, *bacteriophage lambda.* The repressor shuts off the viral genes that code for the protein components of new virus particles and thereby enables the viral genome to remain a silent passenger in the bacterial chromosome, multiplying with the bacterium when conditions are favorable for bacterial growth (see Figure 6–80). The lambda repressor was among the first gene regulatory proteins to be characterized, and it remains one of the best understood, as we discuss later. Other bacterial regulators respond to nutritional conditions by shutting off genes encoding specific sets of metabolic enzymes when they are not needed. The *lac repressor,* for example, the first of these bacterial proteins to be recognized, turns off the production of the proteins responsible for lactose metabolism when this sugar is absent from the medium.

The first step toward understanding gene regulation was the isolation of mutant strains of bacteria and bacteriophage lambda that were unable to shut off specific sets of genes. It was proposed at the time, and later proved, that most of these mutants were deficient in proteins acting as specific repressors for these sets of genes. Because these proteins, like most gene regulatory proteins, are present in small quantities, it was difficult and time-consuming to isolate them. They were eventually purified by fractionating cell extracts on a series of standard chromatography columns (see pp. 166–169). Once isolated, the proteins were shown to bind to specific DNA sequences close to the genes that they

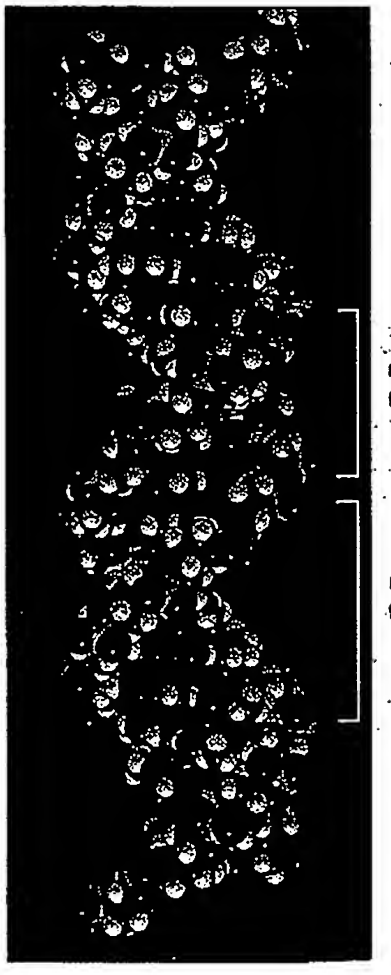


Figure 9–3 Double-helical structure of DNA. The major and minor grooves on the outside of the double helix are indicated. The atoms are colored as follows: carbon, *dark blue;* nitrogen, *light blue;* hydrogen, *white;* oxygen, *red;* phosphorus, *yellow.*

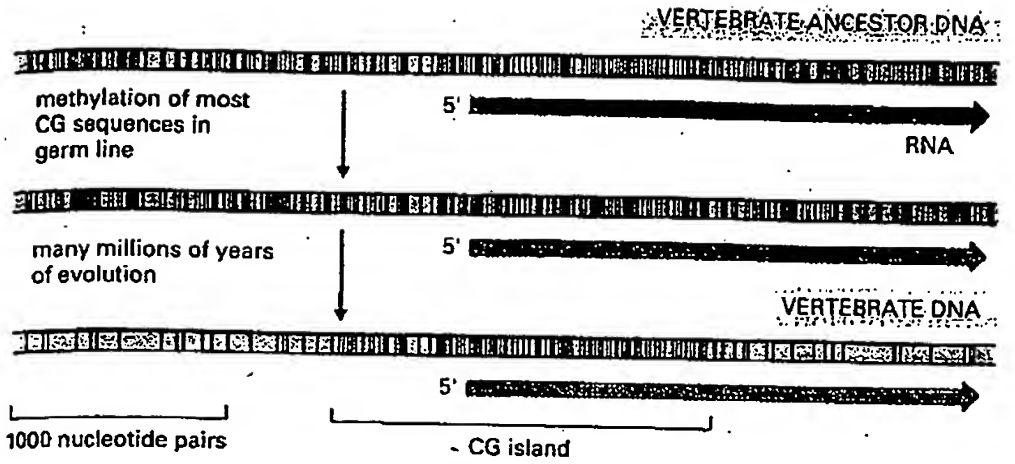


Figure 9–71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A *black line* marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a *red line* marks the location of a methylated CG dinucleotide.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Procaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Posttranscriptional Controls

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls,** which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control,* for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9–72).

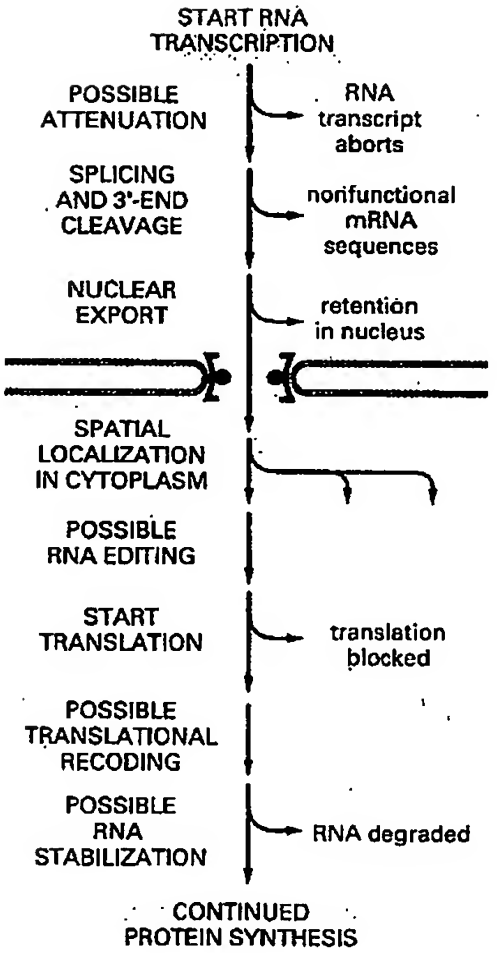


Figure 9–72 Possible posttranscriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.

MOLECULAR BIOLOGY OF THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

Garland Science
Taylor & Francis Group

Garland
Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilernia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

Library of Congress Cataloging-in-Publicaton Data
Molecular biology of the cell / Bruce Alberts ... [et al.].-- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group, 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Cell Biology Interactive
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Front cover Human Genome: Reprinted by permission from *Nature,* International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila, Arabidopsis,* Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

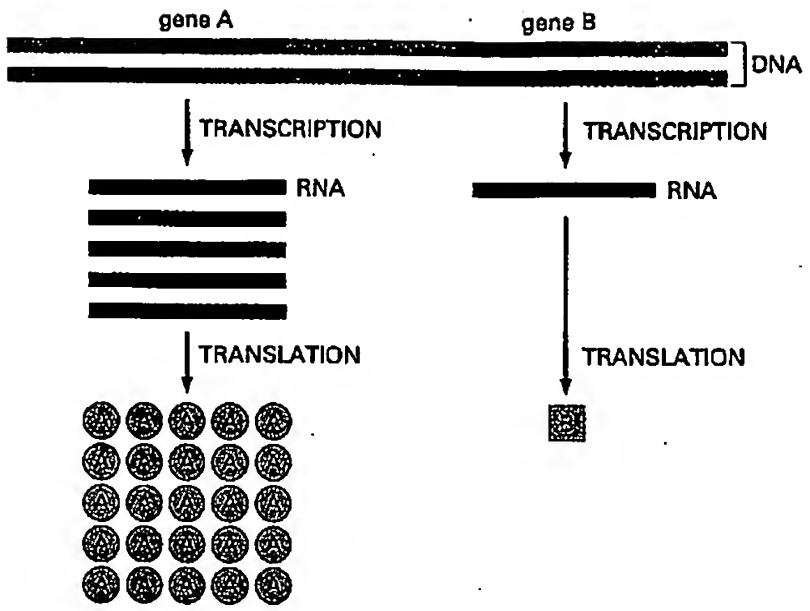


Figure 6–3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6–3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription.**

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6–4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribo*nucleic acid) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6–5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6–6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

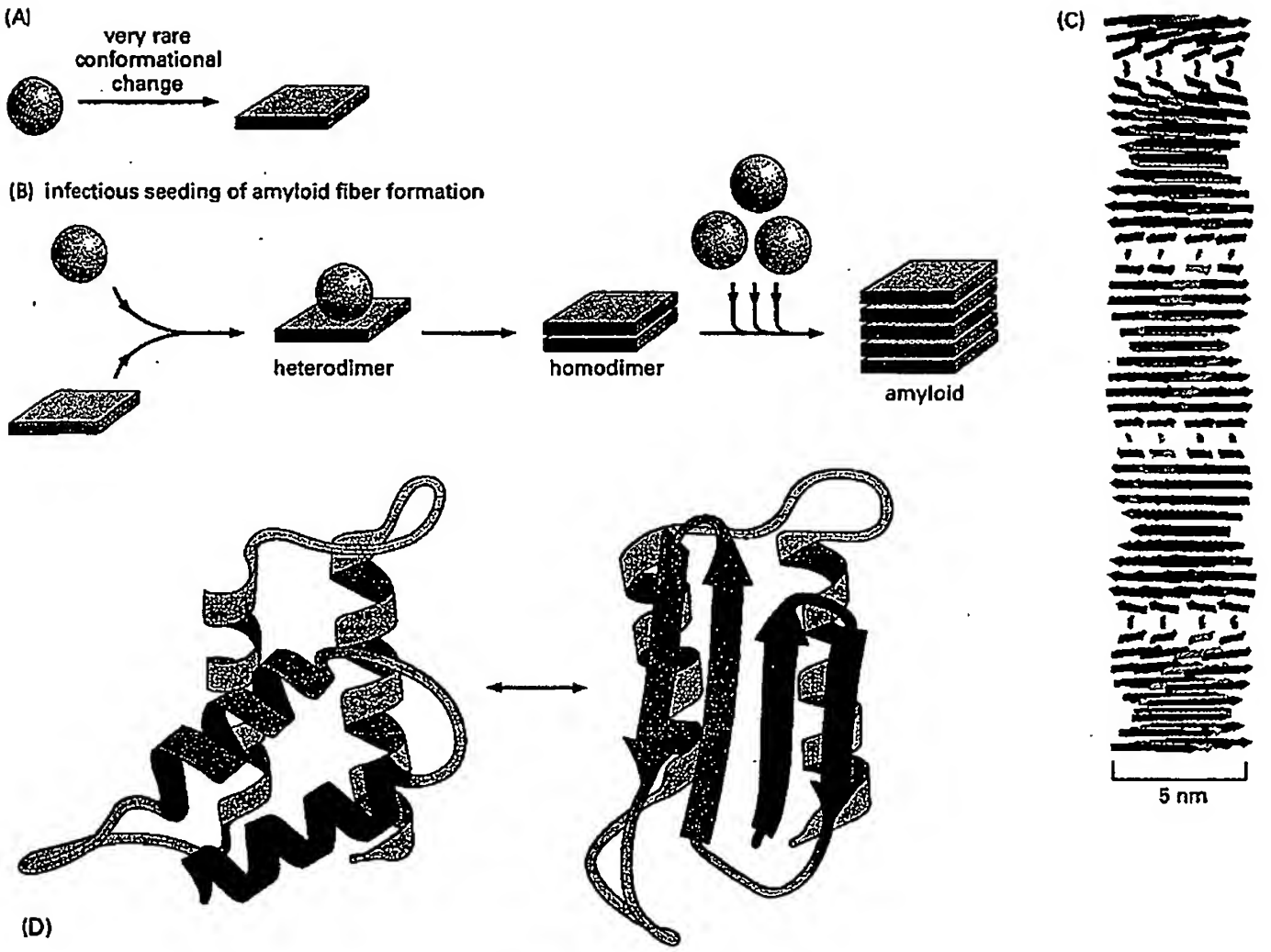


Figure 6–89 Protein aggregates that cause human disease. (A) Schematic illustration of the type of conformational change in a protein that produces material for a cross-beta filament. (B) Diagram illustrating the self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP*, induces the normal PrP protein it contacts to change its conformation, as shown. Most of the human diseases caused by protein aggregation are caused by the overproduction of a variant protein that is especially prone to aggregation, but because this structure is not infectious in this way, it cannot spread from one animal to another. (C) Drawing of a cross-beta filament, a common type of protease-resistant protein aggregate found in a variety of human neurological diseases. Because the hydrogen-bond interactions in a β sheet form between polypeptide backbone atoms (see Figure 3–9), a number of different abnormally folded proteins can produce this structure. (D) One of several possible models for the conversion of PrP to PrP*, showing the likely change of two α-helices into four β-strands. Although the structure of the normal protein has been determined accurately, the structure of the infectious form is not yet known with certainty because the aggregation has prevented the use of standard structural techniques. (C, courtesy of Louise Serpell, adapted from M. Sunde et al., *J. Mol. Biol.* 273:729–739, 1997; D, adapted from S.B. Prusiner, *Trends Biochem. Sci.* 21:482–487, 1996.)

animals and humans. It can be dangerous to eat the tissues of animals that contain PrP*, as witnessed most recently by the spread of BSE (commonly referred to as the "mad cow disease") from cattle to humans in Great Britain.

Fortunately, in the absence of PrP*, PrP is extraordinarily difficult to convert to its abnormal form. Although very few proteins have the potential to misfold into an infectious conformation, a similar transformation has been discovered to be the cause of an otherwise mysterious "protein-only inheritance" observed in yeast cells.

There Are Many Steps From DNA to Protein

We have seen so far in this chapter that many different types of chemical reactions are required to produce a properly folded protein from the information contained in a gene (Figure 6–90). The final level of a properly folded protein in a cell therefore depends upon the efficiency with which each of the many steps is performed.

We discuss in Chapter 7 that cells have the ability to change the levels of their proteins according to their needs. In principle, any or all of the steps in Fig-

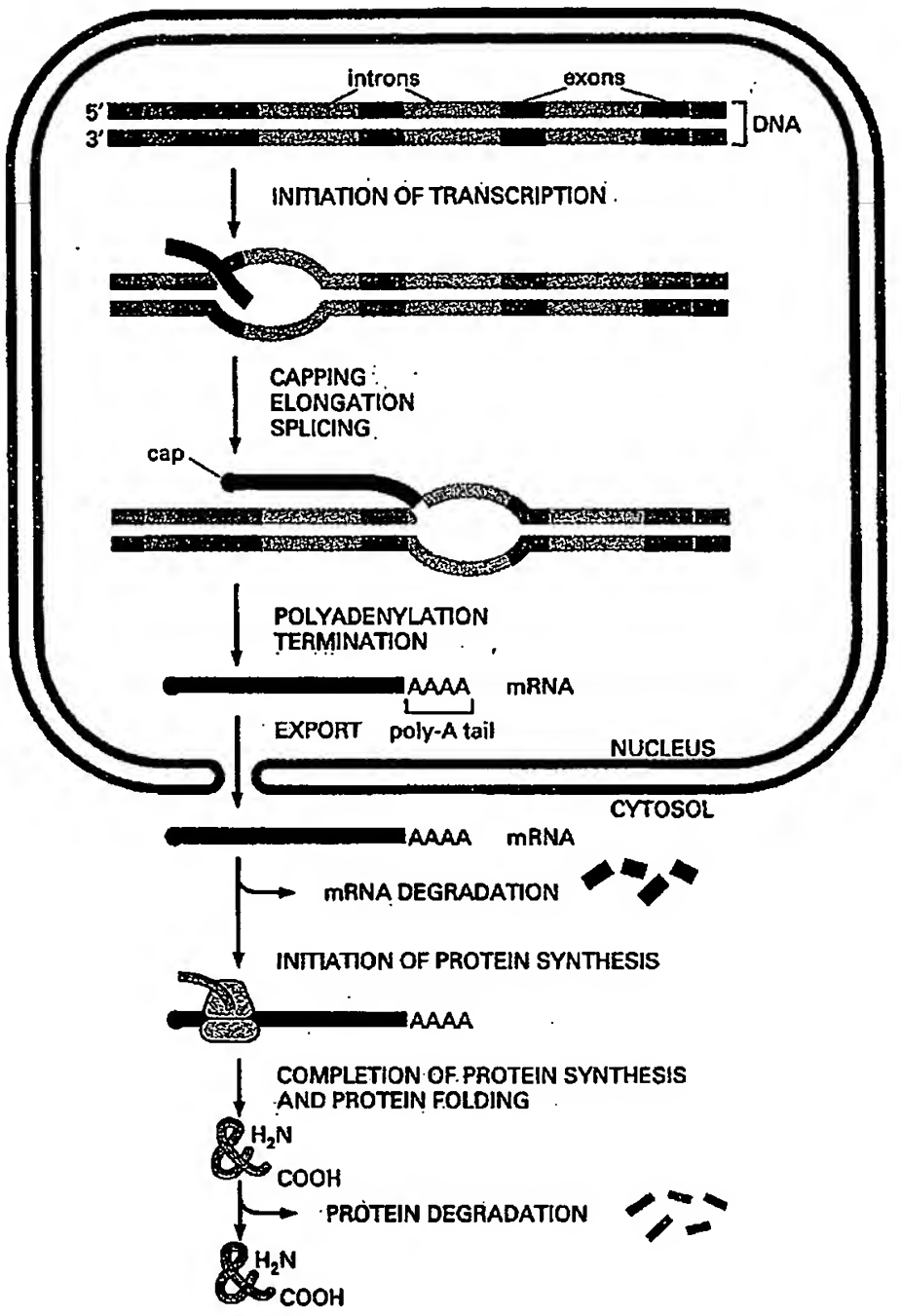


Figure 6–90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6–90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

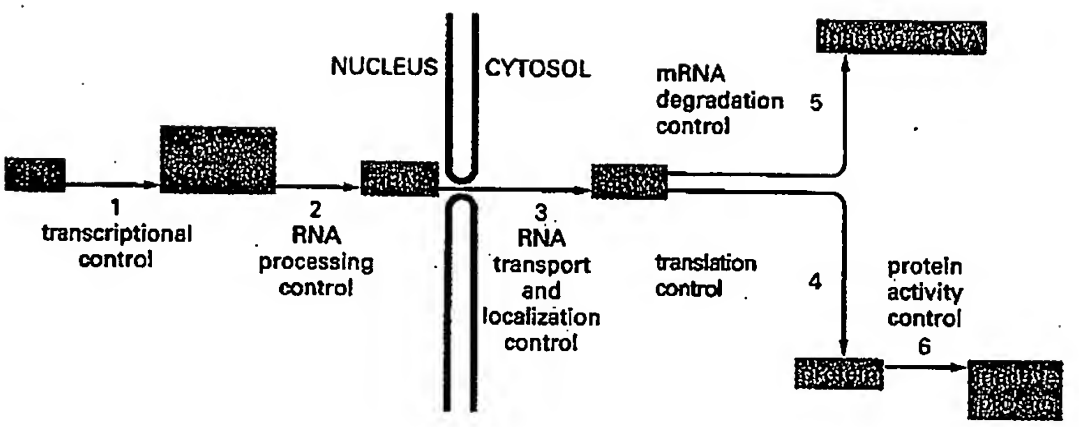


Figure 7–5 Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7–5).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7–5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices

occur in the germ line, the cell lineage that gives rise to sperm or eggs. Most of the DNA in vertebrate germ cells is inactive and highly methylated. Over long periods of evolutionary time, the methylated CG sequences in these inactive regions have presumably been lost through spontaneous deamination events that were not properly repaired. However promoters of genes that remain active in the germ cell lineages (including most housekeeping genes) are kept unmethylated, and therefore spontaneous deaminations of Cs that occur within them can be accurately repaired. Such regions are preserved in modern day vertebrate cells as CG islands. In addition, any mutation of a CG sequence in the genome that destroyed the function or regulation of a gene in the adult would be selected against, and some CG islands are simply the result of a higher than normal density of critical CG sequences.

The mammalian genome contains an estimated 20,000 CG islands. Most of the islands mark the 5′ ends of transcription units and thus, presumably, of genes. The presence of CG islands often provides a convenient way of identifying genes in the DNA sequences of vertebrate genomes.

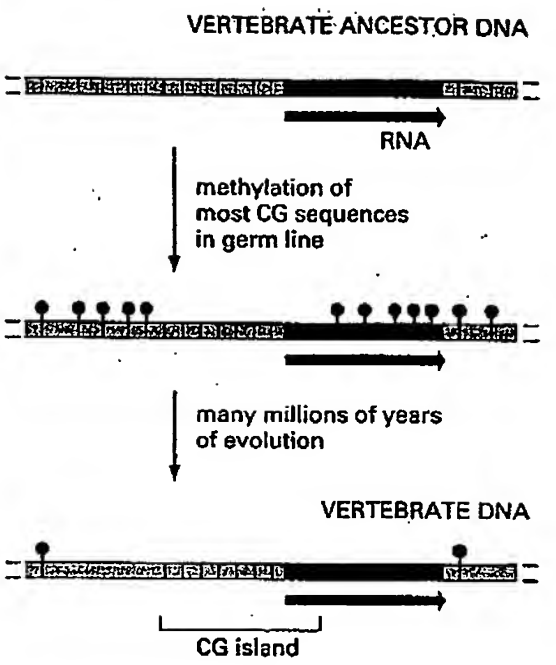


Figure 7–86 A mechanism to explain both the marked overall deficiency of CG sequences and their clustering into CG islands in vertebrate genomes. A *black line* marks the location of a CG dinucleotide in the DNA sequence, while a *red* "lollipop" indicates the presence of a methyl group on the CG dinucleotide. CG sequences that lie in regulatory sequences of genes that are transcribed in germ cells are unmethylated and therefore tend to be retained in evolution. Methylated CG sequences, on the other hand, tend to be lost through deamination of 5-methyl C to T, unless the CG sequence is critical for survival.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character through many cell division cycles and even when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides. These features endow the cell with a memory of its developmental history. Bacteria and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms. One such mechanism involves a competitive interaction between two gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory. Negative feedback loops with programmed delays form the basis for cellular clocks.

In eucaryotes the transcription of a gene is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be active in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also used by eucaryotic cells to regulate gene expression. An especially dramatic case is the inactivation of an entire X chromosome in female mammals. In vertebrates DNA methylation also functions in gene regulation, being used mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms. DNA methylation also underlies the phenomenon of genomic imprinting in mammals, in which the expression of a gene depends on whether it was inherited from the mother or the father.

POST-TRANSCRIPTIONAL CONTROLS

In principle, every step required for the process of gene expression could be controlled. Indeed, one can find examples of each type of regulation, although any one gene is likely to use only a few of them. Controls on the initiation of gene transcription are the predominant form of regulation for most genes. But other controls can act later in the pathway from DNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls**, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control*, for many genes they are crucial.

CHAPTER 29

Regulation of transcription

Genes VII (1997) CH29, pp. 847-848.
Benjamin Lewin

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure
↓
Initiation of transcription
↓
Processing the transcript
↓
Transport to cytoplasm
↓
Translation of mRNA

The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point: probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5′ end, and usually also by polyadenylation at the 3′ end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing: some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein; it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter element that is recognized by a regulatory transcription factor.* An element that causes a gene to respond to such a factor is called a **response element**; examples are the HSE (heat shock response element), GRE (glucocorticoid response element), SRE (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Table 29.1 Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCCNNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. *A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.*

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alter-

World Journal of Surgical Oncology

BioMed Central

Research

Open Access

Prostate stem cell antigen (PSCA) expression in human prostate cancer tissues and its potential role in prostate carcinogenesis and progression of prostate cancer

Zhao Zhigang*[1] and Shen Wenlv[2]

Address: [1]Department of Urology, Shantou University Medical College, Shantou, Guangdong, China and [2]Department of Urology, No 2. Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

Email: Zhao Zhigang* - zgzhaodr@yahoo.com.cn; Shen Wenlv - wlshen99@hotmail.com

* Corresponding author

Published: 10 May 2004

World Journal of Surgical Oncology 2004, 2:13

Received: 30 March 2004

Accepted: 10 May 2004

This article is available from: http://www.wjso.com/content/2/1/13

Abstract

Background: Prostate stem cell antigen (PSCA) is a recently defined homologue of the Thy-1/Ly-6 family of glycosylphosphatidylinositol (GPI)-anchored cell surface antigens. The purpose of the present study was to examine the expression status of PSCA protein and mRNA in clinical specimens of human prostate cancer (Pca) and to validate it as a potential molecular target for diagnosis and treatment of Pca.

Materials and Methods: Immunohistochemical (IHC) and *in situ* hybridization (ISH) analyses of PSCA expression were simultaneously performed on paraffin-embedded sections from 20 benign prostatic hyperplasia (BPH), 20 prostatic intraepithelial neoplasm (PIN) and 48 prostate cancer (Pca) tissues, including 9 androgen-independent prostate cancers. The level of PSCA expression was semiquantitatively scored by assessing both the percentage and intensity of PSCA-positive staining cells in the specimens. Then compared PSCA expression between BPH, PIN and Pca tissues and analysed the correlations of PSCA expression level with pathological grade, clinical stage and progression to androgen-independence in Pca.

Results: In BPH and low grade PIN, PSCA protein and mRNA staining were weak or negative and less intense and uniform than that seen in HGPIN and Pca. There were moderate to strong PSCA protein and mRNA expression in 8 of 11 (72.7%) HGPIN and in 40 of 48 (83.4%) Pca specimens examined by IHC and ISH analyses, with statistical significance compared with BPH (20%) and low grade PIN (22.2%) samples ($p < 0.05$, respectively). The expression level of PSCA increased with high Gleason grade, advanced stage and progression to androgen-independence ($p < 0.05$, respectively). In addition, IHC and ISH staining showed a high degree of correlation between PSCA protein and mRNA overexpression.

Conclusions: Our data demonstrate that PSCA as a new cell surface marker is overexpressed by a majority of human Pca. PSCA expression correlates positively with adverse tumor characteristics, such as increasing pathological grade (poor cell differentiation), worsening clinical stage and androgen-independence, and speculatively with prostate carcinogenesis. PSCA protein overexpression results from upregulated transcription of PSCA mRNA. PSCA may have prognostic utility and may be a promising molecular target for diagnosis and treatment of Pca.

Introduction

Prostate cancer (Pca) is the second leading cause of cancer-related death in American men and is becoming a common cancer increasing in China. Despite recently great progress in the diagnosis and management of localized disease, there continues to be a need for new diagnostic markers that can accurately discriminate between indolent and aggressive variants of Pca. There also continues to be a need for the identification and characterization of potential new therapeutic targets on Pca cells. Current diagnostic and therapeutic modalities for recurrent and metastatic Pca have been limited by a lack of specific target antigens of Pca.

Although a number of prostate-specific genes have been identified (i.e. prostate specific antigen, prostatic acid phosphatase, glandular kallikrein 2), the majority of these are secreted proteins not ideally suited for many immunological strategies. So, the identification of new cell surface antigens is critical to the development of new diagnostic and therapeutic approaches to the management of Pca.

Reiter RE et al [1] reported the identification of prostate stem cell antigen (PSCA), a cell surface antigen that is predominantly prostate specific. The PSCA gene encodes a 123 amino acid glycoprotein, with 30% homology to stem cell antigen 2 (Sca 2). Like Sca-2, PSCA also belongs to a member of the Thy-1/Ly-6 family and is anchored by a glycosylphosphatidylinositol (GPI) linkage. mRNA *in situ* hybridization (ISH) localized PSCA expression in normal prostate to the basal cell epithelium, the putative stem cell compartment of prostatic epithelium, suggesting that PSCA may be a marker of prostate stem/progenitor cells.

In order to examine the status of PSCA protein and mRNA expression in human Pca and validate it as a potential diagnostic and therapeutic target for Pca, we used immunohistochemistry (IHC) and *in situ* hybridization (ISH) simultaneously, and conducted PSCA protein and mRNA expression analyses in paraffin-embedded tissue specimens of benign prostatic hyperplasia (BPH, n = 20), prostate intraepithelial neoplasm (PIN, n = 20) and prostate cancer (Pca, n = 48). Furthermore, we evaluated the possible correlation of PSCA expression level with Pca tumorigenesis, grade, stage and progression to androgen-independence.

Materials and methods

Tissue samples

All of the clinical tissue specimens studied herein were obtained from 80 patients of 57–84 years old by prostatectomy, transurethral resection of prostate (TURP) or biopsies. The patients were classified as 20 cases of BPH, 20 cases of PIN, 40 cases of primary Pca, including 9 patients with recurrent Pca and a history of androgen ablation therapy (orchiectomy and/or hormonal therapy), who were referred to as androgen-independent prostate cancers. Eight specimens were harvested from these androgen-independent Pca patients prior to androgen ablation treatment. Each tissue sample was cut into two parts, one was fixed in 10% formalin for IHC and the other treated with 4% paraformaldehyde/0.1 M PBS PH 7.4 in 0.1% DEPC for 1 h for ISH analysis, and then embedded in paraffin. All paraffin blocks examined were then cut into 5 μm sections and mounted on the glass slides specific for IHC and ISH respectively in the usual fashion. H&E-stained section of each Pca was evaluated and assigned a Gleason score by the experienced urological pathologist at our institution based on the criteria of Gleason score [2]. The Gleason sums are summarized in Table 1. Clinical staging was performed according to Jewett-whitmore-prout staging system, as shown in Table 2. In the category of PIN, we graded the specimens into two groups, i.e. low grade PIN (grade I – II) and high grade PIN (HGPIN, grade III) on the basis of literatures [3,4].

Immunohistochemical (IHC) analysis

Briefly, tissue sections were deparaffinized, dehydrated, and subjected to microwaving in 10 mmol/L citrate buffer, PH 6.0 (Boshide, Wuhan, China) in a 900 W oven for 5 min to induce epitope retrieval. Slides were allowed to cool at room temperature for 30 min. A primary mouse antibody specific to human PSCA (Boshide, Wuhan, China) with a 1:100 dilution was applied to incubate with the slides at room temperature for 2 h. Labeling was detected by sequentially adding biotinylated secondary antibodies and strepavidin-peroxidase, and localized using 3,3'-diaminobenzidine reaction. Sections were then counterstained with hematoxylin. Substitution of the primary antibody with phosphate-buffered-saline (PBS) served as a negative-staining control.

mRNA in situ hybridization (ISH)

Five-μm-thick tissue sections were deparaffinized and dehydrated, then digested in pepsin solution (4 mg/ml in 3% citric acid) for 20 min at 37.5°C, and further processed for ISH. Digoxigenin-labeled sense and antisense human PSCA RNA probes (obtained from Boshide, Wuhan, China) were hybridized to the sections at 48°C overnight. The posthybridization wash with a high stringency was performed sequentially at 37°C in 2 × standard saline citrate (SSC) for 10 min, in 0.5 × SSC for 15 min and in 0.2 × SSC for 30 min. The slides were then incubated to biotinylated mouse anti-digoxigenin antibody at 37.5°C for 1 h followed by washing in 1 × PBS for 20 min at room temperature, and then to strepavidin-peroxidase at 37.5°C for 20 min followed by washing in 1 × PBS for 15 min at room temperature. Subsequently, the slides were developed with diaminobenzidine and then coun-

Table 1: Correlation of PSCA expression with Gleason score

	Intensity × frequency	
Gleason score	0–6 (%)	9 (%)
2–4	5 (83)	1 (17)
5–7	19 (79)	5 (21)
8–10	5 (28)	13 (72)

Table 2: Correlation of PSCA expression with clinical stage

	Intensity × frequency	
Tumor stage	0–6 (%)	9 (%)
≤B	27 (67.5)	13 (32.5)
≥C	2 (25)	6 (75)

terstained with hematoxylin to localize the hybridization signals. Sections hybridized with the sense control probes routinely did not show any specific hybridization signal above background. All slides were hybridized with PBS to substitute for the probes as a negative control.

Scoring methods

To determine the correlation between the results of PSCA immunostaining and mRNA *in situ* hybridization, the same scoring manners are taken in the present study for PSCA protein staining by IHC and PSCA mRNA staining by ISH. Each slide was read and scored by two independently experienced urological pathologists using Olympus BX-41 light microscopes. The evaluation was done in a blinded fashion. For each section, five areas of similar grade were analyzed semiquantitatively for the fraction of cells staining. Fifty percent of specimens were randomly chosen and rescored to determine the degree of interobserver and intraobserver concordance. There was greater than 95% intra- and interobserver agreement.

The intensity of PSCA expression evaluated microscopically was graded on a scale of 0 to 3+ with 3 being the highest expression observed (0, no staining; 1+, mildly intense; 2+, moderately intense; 3+, severely intense). The staining density was quantified as the percentage of cells staining positive for PSCA with the primary antibody or hybridization probe, as follows: 0 = no staining; 1 = positive staining in <25% of the sample; 2 = positive staining in 25%–50% of the sample; 3 = positive staining in >50% of the sample. Intensity score (0 to 3+) was multiplied by the density score (0–3) to give an overall score of 0–9 [1,5]. In this way, we were able to differentiate specimens that may have had focal areas of increased staining from those that had diffuse areas of increased staining [6]. The overall score for each specimen was then categorically assigned to one of the following groups: 0 score, negative expression; 1–2 scores, weak expression; 3–6 scores, moderate expression; 9 score, strong expression.

Statistical analysis

Intensity and density of PSCA protein and mRNA expression in BPH, PIN and Pca tissues were compared using the Chi-square and Student's *t*-test. Univariate associations between PSCA expression and Gleason score, clinical stage and progression to androgen-independence were calculated using Fisher's Exact Test. For all analyses, $p < 0.05$ was considered statistically significant.

Results

PSCA expression in BPH

In general, PSCA protein and mRNA were expressed weakly in individual samples of BPH. Some areas of prostate expressed weak levels (composite score 1–2), whereas other areas were completely negative (composite score 0). Four cases (20%) of BPH had moderate expression of PSCA protein and mRNA (composite score 4–6) by IHC and ISH. In 2/20 (10%) BPH specimens, PSCA mRNA expression was moderate (composite score 3–6), but PSCA protein expression was weak (composite score

2) in one and negative (composite score 0) in the other. PSCA expression was localized to the basal and secretory epithelial cells, and prostatic stroma was almost negative staining for PSCA protein and mRNA in all cases examined.

PSCA expression in PIN

In this study, we detected weak or negative expression of PSCA protein and mRNA (≤2 scores) in 7 of 9 (77.8%) low grade PIN and in 2 of 11 (18.2%) HGPIN, and moderate expression (3–6 scores) in the rest 2 low grade PIN and 5 of 11 (45.5%) HGPIN. One HGPIN with moderate PSCA mRNA expression (6 score) was found weak staining for PSCA protein (2 score) by IHC. Strong PSCA protein and mRNA expression (9 score) were detected in the remaining 3 of 11 (27.3%) HGPIN. There was a statistically significant difference of PSCA protein and mRNA expression levels observed between HGPIN and BPH ($p < 0.05$), but no statistical difference reached between low grade PIN and BPH ($p > 0.05$).

PSCA expression in Pca

In order to determine if PSCA protein and mRNA can be detected in prostate cancers and if PSCA expression levels are increased in malignant compared with benign glands, Forty-eight paraffin-embedded Pca specimens were analysed by IHC and ISH. It was shown that 19 of 48 (39.6%) Pca samples stained very strongly for PSCA protein and mRNA with a score of 9 and another 21 (43.8%) specimens displayed moderate staining with scores of 4–6 (Figure 1). In addition, 4 specimens with moderate to strong PSCA mRNA expression (scores of 4–9) had weak protein staining (a score of 2) by IHC analyses. Overall, Pca expressed a significantly higher level of PSCA protein and mRNA than any other specimen category in this study ($p < 0.05$, compared with BPH and PIN respectively). The result demonstrates that PSCA protein and mRNA are overexpressed by a majority of human Pca.

Correlation of PSCA expression with Gleason score in Pca

Using the semi-quantitative scoring method as described in Materials and Methods, we compared the expression level of PSCA protein and mRNA with Gleason grade of Pca, as shown in Table 1. Prostate adenocarcinomas were graded by Gleason score as 2–4 scores = well-differentiation, 5–7 scores = moderate-differentiation and 8–10 scores = poor-differentiation [7]. Seventy-two percent of Gleason scores 8–10 prostate cancers had very strong staining of PSCA compared to 21% with Gleason scores 5–7 and 17% with 2–4 respectively, demonstrating that poorly differentiated Pca had significantly stronger expression of PSCA protein and mRNA than moderately and well differentiated tumors ($p < 0.05$). As depicted in Figure 1, IHC and ISH analyses showed that PSCA protein and mRNA expression in several cases of poorly differentiated Pca were particularly prominent, with more intense and uniform staining. The results indicate that PSCA expression increases significantly with higher tumor grade in human Pca.

Correlation of PSCA expression with clinical stage in Pca

With regards to PSCA expression in every stage of Pca, we showed the results in Table 2. Seventy-five percent of locally advanced and node positive cancers (i.e. C-D stages) expressed statistically high levels of PSCA versus 32.5% that were organ confined (i.e. A-B stages) ($p < 0.05$). The data demonstrate that PSCA expression increases significantly with advanced tumor stage in human Pca.

Correlation of PSCA expression with androgen-independent progression of Pca

All 9 specimens of androgen-independent prostate cancers stained positive for PSCA protein and mRNA. Eight specimens were obtained from patients managed prior to androgen ablation therapy. Seven of eight (87.5%) of these androgen-independent prostate cancers were in the strongest staining category (score = 9), compared with three out of eight (37.5%) of patients with androgen-dependent cancers ($p < 0.05$). The results demonstrate that PSCA expression increases significantly with progression to androgen-independence of human Pca.

It is evident from the results above that within a majority of human prostate cancers the level of PSCA protein and mRNA expression correlates significantly with increasing grade, worsening stage and progression to androgen-independence.

Correlation of PSCA immunostaining and mRNA in situ hybridization

In all 88 specimens surveyed herein, we compared the results of PSCA IHC staining with mRNA ISH analysis. Positive staining areas and its intensity and density scores evaluated by IHC were identical to those seen by ISH in 79 of 88 (89.8%) specimens (18/20 BPH, 19/20 PIN and 42/48 Pca respectively). Importantly, 27/27 samples with PSCA mRNA composite scores of 0–2, 32/36 samples with scores of 3–6 and 22/24 samples with a score of 9 also had PSCA protein expression scores of 0–2, 3–6 and 9 respectively. However, in 5 samples with PSCA mRNA overall scores of 3–6 and in 2 with scores of 9 there were less or negative PSCA protein expression (i.e. scores of 0–4), suggesting that this may reflect posttranscriptional modification of PSCA or that the epitopes recognized by PSCA mAb may be obscured in some cancers. The data demonstrate that the results of PSCA immunostaining were consistent with those of mRNA ISH analysis, showing a high degree of correlation between PSCA protein and mRNA expression.

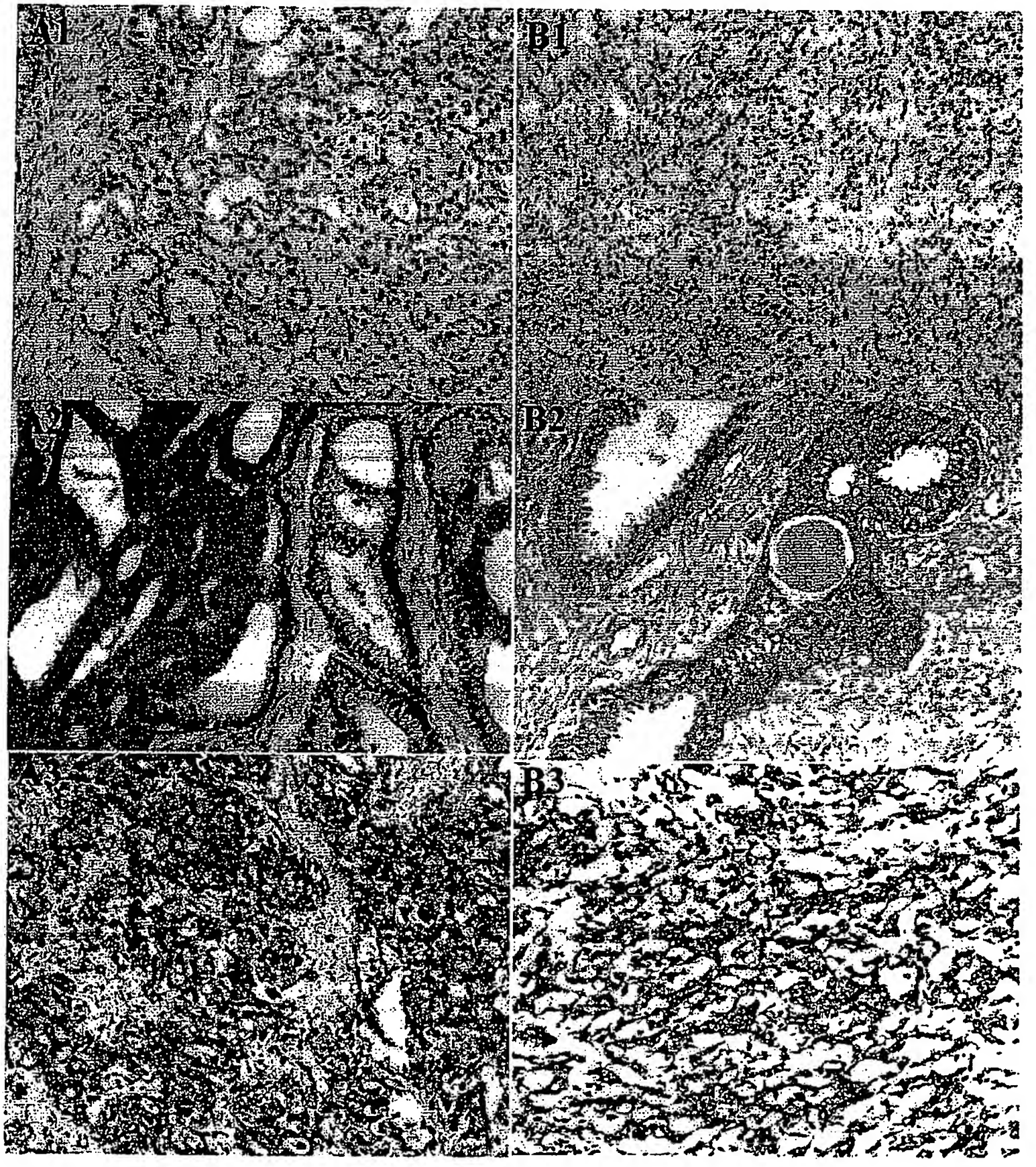


Figure 1

Representatives of PSCA IHC and ISH staining in Pca (A. IHC staining, B. ISH staining, ×200 magnification). A_1, B_1: negative control of IHC and ISH. PBS replacing the primary antibody (A_1) and hybridization with a sense PSCA probe (B_1) showed no background staining. A_2, B_2: a moderately differentiated Pca (Gleason score = 3+3 = 6) with moderate staining (composite score = 6) in all malignant cells; A_2: IHC shows not only cell surface but also apparent cytoplasmic staining of PSCA protein. A_3, B_3: a poorly differentiated Pca (Gleason score = 4+4 = 8) with very strong staining (composite score = 9) in all malignant cells.

Discussion

PSCA is homologous to a group of cell surface proteins that mark the earliest phase of hematopoietic development. PSCA mRNA expression is prostate-specific in normal male tissues and is highly up-regulated in both androgen-dependent and-independent Pca xenografts (LAPC-4 tumors). We hypothesize that PSCA may play a role in Pca tumorigenesis and progression, and may serve as a target for Pca diagnosis and treatment. In this study, IHC and ISH showed that in general there were weak or absent PSCA protein and mRNA expression in BPH and low grade PIN tissues. However, PSCA protein and mRNA are widely expressed in HGPIN, the putative precursor of invasive Pca, suggesting that up-regulation of PSCA is an early event in prostate carcinogenesis. Recently, Reiter RE et al [1], using ISH analysis, reported that 97 of 118 (82%) HGPIN specimens stained strongly positive for PSCA mRNA. A very similar finding was seen on mouse PSCA (mPSCA) expression in mouse HGPIN tissues by Tran C. P et al [8]. These data suggest that PSCA may be a new marker associated with transformation of prostate cells and tumorigenesis.

We observed that PSCA protein and mRNA are highly expressed in a large percentage of human prostate cancers, including advanced, poorly differentiated, androgen-independent and metastatic cases. Fluorescence-activated cell sorting and confocal/ immunofluorescent studies demonstrated cell surface expression of PSCA protein in Pca cells [9]. Our IHC expression analysis of PSCA shows not only cell surface but also apparent cytoplasmic staining of PSCA protein in Pca specimens (Figure 1). One possible explanation for this is that anti-PSCA antibody can recognize PSCA peptide precursors that reside in the cytoplasm. Also, it is possible that the positive staining that appears in the cytoplasm is actually from the overlying cell membrane [5]. These data seem to indicate that PSCA is a novel cell surface marker for human Pca.

Our results show that elevated level of PSCA expression correlates with high grade (i.e. poor differentiation), increased tumor stage and progression to androgen-independence of Pca. These findings support the original IHC analyses by Gu Z et al [9], who reported that PSCA protein expressed in 94% of primary Pca and the intensity of PSCA protein expression increased with tumor grade, stage and progression to androgen-independence. Our results also collaborate the recent work of Han KR et al [10], in which the significant association between high PSCA expression and adverse prognostic features such as high Gleason score, seminal vesicle invasion and capsular involvement in Pca was found. It is suggested that PSCA overexpression may be an adverse predictor for recurrence, clinical progression or survival of Pca. Hara H et al [11] used RT-PCR detection of PSA, PSMA and PSCA in 1 ml of peripheral blood to evaluate Pca patients with poor prognosis. The results showed that among 58 PCa patients, each PCR indicated the prognostic value in the hierarchy of PSCA>PSA>PSMA RT-PCR, and extraprostatic cases with positive PSCA PCR indicated lower disease-progression-free survival than those with negative PSCA PCR, demonstrating that PSCA can be used as a prognostic factor. Dubey P et al [12] reported that elevated numbers of PSCA + cells correlate positively with the onset and development of prostate carcinoma over a long time span in the prostates of the TRAMP and PTEN +/- models compared with its normal prostates. Taken together with our present findings, in which PSCA is overexpressed from HGPIN to almost frank carcinoma, it is reasonable and possible to use increased PSCA expression level or increased numbers of PSCA-positive cells in the prostate samples as a prognostic marker to predict the potential onset of this cancer. These data raise the possibility that PSCA may have diagnostic utility or clinical prognostic value in human Pca.

The cause of PSCA overexpression in Pca is not known. One possible mechanism is that it may result from PSCA gene amplification. In humans, PSCA is located on chromosome 8q24.2 [1], which is often amplified in metastatic and recurrent Pca and considered to indicate a poor prognosis [13-15]. Interestingly, PSCA is in close proximity to the c-myc oncogene, which is amplified in >20% of recurrent and metastatic prostate cancers [16,17]. Reiter RE et al [18] reported that PSCA and MYC gene copy numbers were co-amplified in 25% of tumors (five out of twenty), demonstrating that PSCA overexpression is associated with PSCA and MYC coamplification in Pca. Gu Z et al [9] recently reporteted that in 102 specimens available to compare the results of PSCA immunostaining with their previous mRNA ISH analysis, 92 (90.2%) had identically positive areas of PSCA protein and mRNA expression. Taken together with our findings, in which we detected moderate to strong expression of PSCA protein and mRNA in 34 of 40 (85%) Pca specimens examined simultaneously by IHC and ISH analyses, it is demonstrated that PSCA protein and mRNA overexpressed in human Pca, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA.

At present, the regulation mechanisms of human PSCA expression and its biological function are yet to be elucidated. PSCA expression may be regulated by multiple factors [18]. Watabe T et al [19] reported that transcriptional control is a major component regulating PSCA expression levels. In addition, induction of PSCA expression may be regulated or mediated through cell-cell contact and protein kinase C (PKC) [20]. Homologues of PSCA have diverse activities, and have themselves been involved in

carcinogenesis. Signalling through SCA-2 has been demonstrated to prevent apoptosis in immature thymocytes [21]. Thy-1 is involved in T cell activation and transducts signals through src-like tyrosine kinases [22]. Ly-6 genes have been implicated both in tumorigenesis and in cell-cell adhesion [23-25]. Cell-cell or cell-matrix interaction is critical for local tumor growth and spread to distal sites. From its restricted expression in basal cells of normal prostate and its homology to SCA-2, PSCA may play a role in stem/progenitor cell function, such as self-renewal (i.e. anti-apoptosis) and/or proliferation [1]. Taken together with the results in the present study, we speculate that PSCA may play a role in tumorigenesis and clinical progression of Pca through affecting cell transformation and proliferation. From our results, it is also suggested that PSCA as a new cell surface antigen may have a number of potential uses in the diagnosis, therapy and clinical prognosis of human Pca. PSCA overexpression in prostate biopsies could be used to identify patients at high risk to develop recurrent or metastatic disease, and to discriminate cancers from normal glands in prostatectomy samples. Similarly, the detection of PSCA-overexpressing cells in bone marrow or peripheral blood may identify and predict metastatic progression better than current assays, which identify only PSA-positive or PSMA-positive prostate cells.

In summary, we have shown in this study that PSCA protein and mRNA are maintained in expression from HGPIN through all stages of Pca in a majority of cases, which may be associated with prostate carcinogenesis and correlate positively with high tumor grade (poor cell differentiation), advanced stage and androgen-independent progression. PSCA protein overexpression is due to the upregulation of its mRNA transcription. The results suggest that PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor.

Competing interests

None declared.

References

1. Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, David E, Wahl M, Nisitani S, Yamashiro J, Le Beau MM, Loda M, Witte ON: **Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer.** *Proc Natl Acad Sci USA* 1998, 95:1735-1740.
2. Gleason DF: **Histologic grading and clinical staging of prostatic carcinoma.** In: *Urologic Pathology: The Prostate* Edited by: Tannebaum M. Philadelphia, Lea & Febiger; 1977:171-197.
3. Brawer MK: **Prostatic intraepithelial neoplasia: a premalignant lesion.** *Hum Pathol* 1992, 23:242-248.
4. Amin MB, Ro JY, Ayala AC: **Prostatic intraepithelial neoplasia: relationship to adenocarcinoma of prostate.** *Pathol Annu* 1994, 29:1-30.
5. Amara N, Palapattu GS, Schrage M, Gu Z, Thomas GV, Dorey F, Said J, Reiter RE: **Prostate stem cell antigen is overexpressed in human transitional cell carcinoma.** *Cancer Res* 2001, 61:4660-4665.
6. Hanas JS, Lerner MR, Lightfoot SA, Raczkowski C, Kastens DJ, Brackett DJ, Postier RG: **Expression of the cyclin-dependent kinase inhibitor p21 (WAF1/CIP1) and p53 tumor suppressor in dysplastic progression and adenocarcinoma in Barrett esophagus.** *Cancer (Phila)* 1999, 86:756-763.
7. Egevad L, Gramfors T, Karlberg L: **Prognostic value of the Gleason score in prostate cancer.** *BJU Int* 2002, 89:538-542.
8. Tran CP, Lin C, Yamashiro J, Reiter RE: **Prostate stem cell antigen is a marker of late intermediate prostate epithelial cells.** *Mol Cancer Res* 2002, 1:113-121.
9. Gu Z, Thomas G, Yamashiro J, Shintaku IP, Dorey F, Raitano A, Witte ON, Said JW, Loda M, Reiter RE: **Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer.** *Oncogene* 2000, 19:1288-1296.
10. Han KR, Seligson DB, Liu X, Horvath S, Shintaku PI, Thomas GV, Said JW, Reiter RE: **Prostate stem cell antigen expression is associated with gleason score, seminal vesicle invasion and capsular invasion in prostate cancer.** *J Urol* 2004, 171:1117-1121.
11. Hara H, Kasahara T, Kawasaki T, Bilim V, Obara K, Takahashi K, Tomita Y: **Reverse Transcription-Polymerase Chain Reaction Detection of Prostate-specific Antigen, Prostate-specific Membrane Antigen, and Prostate Stem Cell Antigen in One Milliliter of Peripheral Blood.** *Clin Cancer Res* 2002, 8:1794-1799.
12. Dubey P, Wu H, Reiter RE, Witte ON: **Alternative pathways to prostate carcinoma activate prostate stem cell antigen expression.** *Cancer Res* 2001, 61:3256-3261.
13. Visa korpi T, Kallioniemi AH, Syvanen AC, Hyytinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP: **Genetic changes in primary and recrrent prostate cancer by comparative genomic hybridization.** *Cancer Res* 1995, 55:342-347.
14. Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstralh EJ, Jenkins RB: **Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma.** *J Natl Cancer Inst* 1999, 91:1574-1580.
15. Van Den Berg C, Guan XY, Von Hoff D, Jenkins R, Bittner , Griffin C, Kallioniemi O, Visakorpi , McGill , Herath J, Epstein J, Sarosdy M, Meltzer P, Trent J: **DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications.** *Clin Cancer Res* 1995, 1:11-18.
16. Jenkins RB, Qian J, Lieber MM, Bostwick DG: **Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization.** *Cancer Res* 1997, 57:524-531.
17. Nupponen NN, Kakkola L, Koivisto P, Visakorpi T: **Genetic alterations in hormone-refractory recurrent prostate carcinomas.** *Am J Pathol* 1998, 153:141-148.
18. Reiter RE, Sato I, Thomas G, Qian J, Gu Z, Watabe T, Loda M, Jenkins RB: **Coamplification of prostate stem cell antigen (PSCA) and MYC in locally advanced prostate cancer.** *Genes Chromosomes Cancer* 2000, 27:95-103.
19. Watabe T, Lin M, Donjacour AA, Cunha GR, Witte ON, Reiter RE: **Growth, regeneration, and tumorigenesis of the prostate activates the PSCA promoter.** *Proc Natl Acad Sci USA* 2002, 99:401-406.
20. Bahrenberg G, Brauers A, Joost HG, Jakse G: **PSCA expression is regulated by phorbol ester and cell adhesion in the bladder carcinoma cell line RT112.** *Cancer Lett* 2001, 168:37-43.
21. Noda S, Kosugi A, Saitoh S, Narumiya S, Hamaoka T: **Protection from anti-TCR/CD3-induced apoptosis in immature thymocytes by a signal through thymic shared antigen-1/stem cell antigen-2.** *J Exp Med* 1996, 183:2355-2360.
22. Thomas PM, Samelson LE: **The glycophosphatidylinositol-anchored Thy-1 molecule interacts with the p60fyn protein tyrosine kinase in T cells.** *J Biol Chem* 1992, 267:12317-12322.
23. Bamezai A, Rock KL: **Overexpressed Ly-6A.2 mediated cell-cell adhesion by binding a ligand expressed on lymphoid cells.** *Proc Natl Acad Sci USA* 1995, 92:4294-4298.
24. Katz BZ, Eshel R, Sagi-Assif O, Witz IP: **An association between high Ly-6A/E expression on tumor cells and a highly malignant phenotype.** *Int J Cancer* 1994, 59:684-691.
25. Brakenhoff RH, Gerretsen M, Knippels EM, van Dijk M, van Essen H, Weghuis DO, Sinke RJ, Snow GB, van Dongen GA: **The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion.** *J Cell Biol* 1995, 129:1677-1689.

Review

Translation Initiation in Cancer: A Novel Target for Therapy[1]

Funda Meric[2] and Kelly K. Hunt

Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Abstract

Translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled with cell cycle progression and cell growth. Several alterations in translational control occur in cancer. Variant mRNA sequences can alter the translational efficiency of individual mRNA molecules, which in turn play a role in cancer biology. Changes in the expression or availability of components of the translational machinery and in the activation of translation through signal transduction pathways can lead to more global changes, such as an increase in the overall rate of protein synthesis and translational activation of the mRNA molecules involved in cell growth and proliferation. We review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to help elucidate new therapeutic avenues.

Introduction

The fundamental principle of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells. With the advent of cDNA array technology, most efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable either to DNA amplification or to differences in transcription. Gene expression is quite complicated, however, and is also regulated at the level of mRNA stability, mRNA translation, and protein stability.

The power of translational regulation has been best recognized among developmental biologists, because transcription does not occur in early embryogenesis in eukaryotes. For example, in *Xenopus*, the period of transcriptional quiescence continues until the embryo reaches midblastula transition, the 4000-cell stage. Therefore, all necessary mRNA molecules are transcribed during oogenesis and stockpiled in a translationally inactive, masked form. The mRNA are translationally activated at appropriate times during oocyte maturation, fertilization, and early embryogenesis and thus, are under strict translational control.

Translation has an established role in cell growth. Basically, an increase in protein synthesis occurs as a consequence of mitogenesis. Until recently, however, little was known about the alterations in mRNA translation in cancer, and much is yet to be discovered about their role in the development and progression of cancer. Here we review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to elucidate potential new therapeutic avenues.

Basic Principles of Translational Control

Mechanism of Translation Initiation

Translation initiation is the main step in translational regulation. Translation initiation is a complex process in which the initiator tRNA and the 40S and 60S ribosomal subunits are recruited to the 5′ end of a mRNA molecule and assembled by eukaryotic translation initiation factors into an 80S ribosome at the start codon of the mRNA (Fig. 1). The 5′ end of eukaryotic mRNA is capped, *i.e.*, contains the cap structure m^7GpppN (7-methylguanosine-triphospho-5′-ribonucleoside). Most translation in eukaryotes occurs in a cap-dependent fashion, *i.e.*, the cap is specifically recognized by the eIF4E,[3] which binds the 5′ cap. The eIF4F translation initiation complex is then formed by the assembly of eIF4E, the RNA helicase eIF4A, and eIF4G, a scaffolding protein that mediates the binding of the 40S ribosomal subunit to the mRNA molecule through interaction with the eIF3 protein present on the 40S ribosome. eIF4A and eIF4B participate in melting the secondary structure of the 5′ UTR of the mRNA. The 43S initiation complex (40S/eIF2/Met-tRNA/GTP complex) scans the mRNA in a 5′→3′ direction until it encounters an AUG start codon. This start codon is then base-paired to the anticodon of initiator tRNA, forming the 48S initiation complex. The initiation factors are then displaced from the 48S complex, and the 60S ribosome joins to form the 80S ribosome.

Unlike most eukaryotic translation, translation initiation of certain mRNAs, such as the picornavirus RNA, is cap independent and occurs by internal ribosome entry. This mechanism does not require eIF4E. Either the 43S complex can bind the initiation codon directly through interaction with the IRES in the 5′ UTR such as in the encephalomyocarditis virus, or it can

Received 5/16/02; revised 7/12/02; accepted 7/22/02.
[1] F. M. is supported by The University of Texas M. D. Anderson Cancer Center Physician-Scientist Program and by NIH Grant 1KO8-CA 91895-01. K. K. H. is supported by Department of Defense Award DAMD-17-97-1-7162.
[2] To whom requests for reprints should be addressed, at Department of Surgical Oncology, Box 444, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: (713) 745-4453; Fax: (713) 745-4926; E-mail: fmeric@mdanderson.org.

[3] The abbreviations used are: eIF4E, eukaryotic initiation factor 4E; UTR, untranslated region; IRES, internal ribosome entry site; 4E-BP1, eukaryotic initiation factor 4E-binding protein 1; S6K, ribosomal p70 S6 kinase; mTOR, mammalian target of rapamycin; ATM, ataxia telangiectasia mutated; PI3K, phosphatidylinositol 3-kinase; PTEN, phosphatase and tensin homolog deleted from chromosome 10; PP2A, protein phosphatase 2A; TGF-β3, transforming growth factor-β3; PAP, poly(A) polymerase; EPA, eicosopentaenoic acid; *mda-7*, melanoma differentiation-associated gene 7.

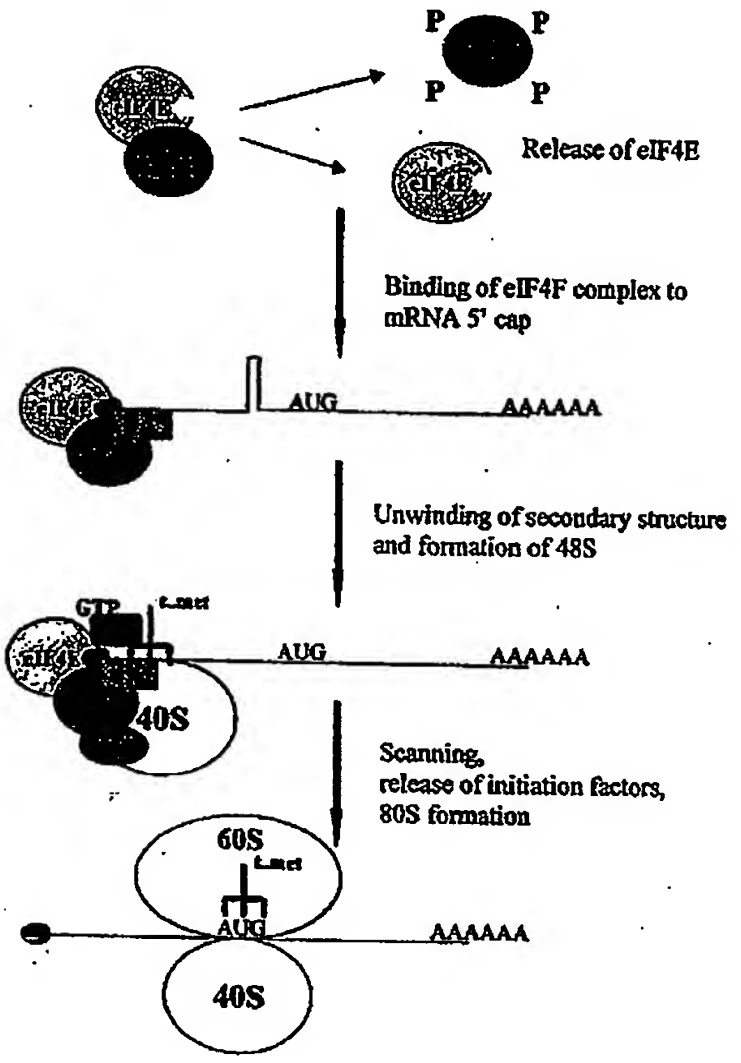


Fig. 1. Translation initiation in eukaryotes. The 4E-BPs are hyperphosphorylated to release eIF4E so that it can interact with the 5′ cap, and the eIF4F initiation complex is assembled. The interaction of poly(A) binding protein with the initiation complex and circularization of the mRNA is not depicted in the diagram. The secondary structure of the 5′ UTR is melted, the 40S ribosomal subunit is bound to eIF3, and the ternary complex consisting of eIF2, GTP, and the Met-tRNA are recruited to the mRNA. The ribosome scans the mRNA in a 5′→3′ direction until an AUG start codon is found in the appropriate sequence context. The initiation factors are released, and the large ribosomal subunit is recruited.

initially attach to the IRES and then reach the initiation codon by scanning or transfer, as is the case with the poliovirus (1).

Regulation of Translation Initiation

Translation initiation can be regulated by alterations in the expression or phosphorylation status of the various factors involved. Key components in translational regulation that may provide potential therapeutic targets follow.

eIF4E. eIF4E plays a central role in translation regulation. It is the least abundant of the initiation factors and is considered the rate-limiting component for initiation of cap-dependent translation. eIF4E may also be involved in mRNA splicing, mRNA 3′ processing, and mRNA nucleocytoplasmic transport (2). eIF4E expression can be increased at the transcriptional level in response to serum or growth factors (3). eIF4E overexpression may cause preferential translation of mRNAs containing excessive secondary structure in their 5′ UTR that are normally discriminated against by the translational machinery and thus are inefficiently translated (4–7). As examples of this, overexpression of eIF4E promotes increased translation of vascular endothelial growth factor, fibroblast growth factor-2, and cyclin D1 (2, 8, 9).

Another mechanism of control is the regulation of eIF4E phosphorylation. eIF4E phosphorylation is mediated by the mitogen-activated protein kinase-interacting kinase 1, which is activated by the mitogen-activated pathway activating extracellular signal-related kinases and the stress-activated pathway acting through p38 mitogen-activated protein kinase (10–13). Several mitogens, such as serum, platelet-derived growth factor, epidermal growth factor, insulin, angiotensin II, src kinase overexpression, and ras overexpression, lead to eIF4E phosphorylation (14). The phosphorylation status of eIF4E is usually correlated with the translational rate and growth status of the cell; however, eIF4E phosphorylation has also been observed in response to some cellular stresses when translational rates actually decrease (15). Thus, further study is needed to understand the effects of eIF4E phosphorylation on eIF4E activity.

Another mechanism of regulation is the alteration of eIF4E availability by the binding of eIF4E to the eIF4E-binding proteins (4E-BP, also known as PHAS-I). 4E-BPs compete with eIF4G for a binding site in eIF4E. The binding of eIF4E to the best characterized eIF4E-binding protein, 4E-BP1, is regulated by 4E-BP1 phosphorylation. Hypophosphorylated 4E-BP1 binds to eIF4E, whereas 4E-BP1 hyperphosphorylation decreases this binding. Insulin, angiotensin, epidermal growth factor, platelet-derived growth factor, hepatocyte growth factor, nerve growth factor, insulin-like growth factors I and II, interleukin 3, granulocyte-macrophage colony-stimulating factor + steel factor, gastrin, and the adenovirus have all been reported to induce phosphorylation of 4E-BP1 and to decrease the ability of 4E-BP1 to bind eIF4E (15, 16). Conversely, deprivation of nutrients or growth factors results in 4E-BP1 dephosphorylation, an increase in eIF4E binding, and a decrease in cap-dependent translation.

p70 S6 Kinase. Phosphorylation of ribosomal 40S protein S6 by S6K is thought to play an important role in translational regulation. S6K −/− mouse embryonic cells proliferate more slowly than do parental cells, demonstrating that S6K has a positive influence on cell proliferation (17). S6K regulates the translation of a group of mRNAs possessing a 5′ terminal oligopyrimidine tract (5′ TOP) found at the 5′ UTR of ribosomal protein mRNAs and other mRNAs coding for components of the translational machinery. Phosphorylation of S6K is regulated in part based on the availability of nutrients (18, 19) and is stimulated by several growth factors, such as platelet-derived growth factor and insulin-like growth factor I (20).

eIF2α Phosphorylation. The binding of the initiator tRNA to the small ribosomal unit is mediated by translation initiation factor eIF2. Phosphorylation of the α-subunit of eIF2 prevents formation of the eIF2/GTP/Met-tRNA complex and inhibits global protein synthesis (21, 22). eIF2α is phosphorylated under a variety of conditions, such as viral infection, nutrient deprivation, heme deprivation, and apoptosis (22). eIF2α is phosphorylated by heme-regulated inhibitor, nutrient-regulated protein kinase, and the IFN-induced, double-stranded RNA-activated protein kinase (PKR; Ref. 23).

The mTOR Signaling Pathway. The macrolide antibiotic rapamycin (Sirolimus; Wyeth-Ayerst Research, Collegeville, PA) has been the subject of intensive study because it inhibits signal transduction pathways involved in T-cell activation. The rapamycin-sensitive component of these pathways is mTOR (also called FRAP or RAFT1). mTOR is the mammalian homologue of the yeast TOR proteins that regulate G_1 progression and translation in response to nutrient availability (24). mTOR is a serine-threonine kinase that modulates translation initiation by altering the phosphorylation status of 4E-BP1 and S6K (Fig. 2; Ref. 25).

4E-BP1 is phosphorylated on multiple residues. mTOR phosphorylates the Thr-37 and Thr-46 residues of 4E-BP1 *in vitro* (26); however, phosphorylation at these sites is not associated with a loss of eIF4E binding. Phosphorylation of Thr-37 and Thr-46 is required for subsequent phosphorylation at several COOH-terminal, serum-sensitive sites; a combination of these phosphorylation events appears to be needed to inhibit the binding of 4E-BP1 to eIF4E (25). The product of the *ATM* gene, p38/MSK1 pathway, and protein kinase Cσ also play a role in 4E-BP1 phosphorylation (27–29).

S6K and 4E-BP1 are also regulated, in part, by PI3K and its downstream protein kinase Akt. PTEN is a phosphatase that negatively regulates PI3K signaling. PTEN null cells have constitutively active of Akt, with increased S6K activity and S6 phosphorylation (30). S6K activity is inhibited both by PI3K inhibitors wortmannin and LY294002 and by mTOR inhibitor rapamycin (24). Akt phosphorylates Ser-2448 in mTOR *in vitro*, and this site is phosphorylated upon Akt activation *in vivo* (31–33). Thus, mTOR is regulated by the PI3K/Akt pathway; however, this does not appear to be the only mode of regulation of mTOR activity. Whether the PI3K pathway also regulates S6K and 4E-BP1 phosphorylation independent of mTOR is controversial.

Interestingly, mTOR autophosphorylation is blocked by wortmannin but not by rapamycin (34). This seeming inconsistency suggests that mTOR-responsive regulation of 4E-BP1 and S6K activity occurs through a mechanism other than intrinsic mTOR kinase activity. An alternate pathway for 4E-BP1 and S6K phosphorylation by mTOR activity is by the inhibition of a phosphatase. Treatment with calyculin A, an inhibitor of phosphatases 1 and 2A, reduces rapamycin-induced dephosphorylation of 4E-BP1 and S6K by rapamycin (35). PP2A interacts with full-length S6K but not with a S6K mutant that is resistant to dephosphorylation resulting from rapamycin. mTOR phosphorylates PP2A *in vitro*; however, how this process alters PP2A activity is not known. These results are consistent with the model that phosphorylation of a phosphatase by mTOR prevents dephosphorylation of 4E-BP1 and S6K, and conversely, that nutrient deprivation and rapamycin block inhibition of the phosphatase by mTOR.

Polyadenylation. The poly(A) tail in eukaryotic mRNA is important in enhancing translation initiation and mRNA stability. Polyadenylation plays a key role in regulating gene expression during oogenesis and early embryogenesis. Some mRNA that are translationally inactive in the oocyte are polyadenylated concomitantly with translational activation in oocyte maturation, whereas other mRNAs that are translationally active during oogenesis are deadenylated and translationally silenced (36–38). Thus, control of poly(A) tail synthesis is an important regulatory step in gene expression. The 5′ cap and poly(A) tail are thought to function synergistically to regulate mRNA translational efficiency (39, 40).

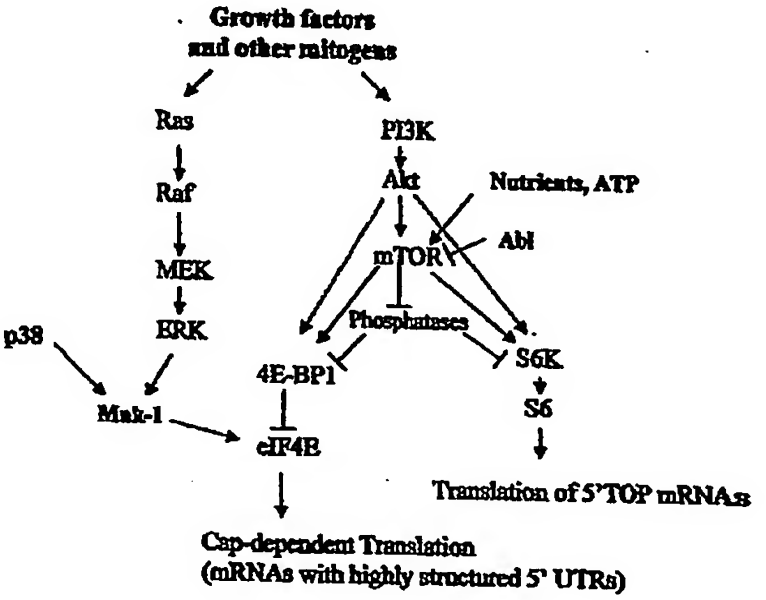


Fig. 2. Regulation of translation initiation by signal transduction pathways. Signaling via p38, extracellular signal-related kinase, PI3K, and mTOR can all activate translation initiation.

RNA Packaging. Most RNA-binding proteins are assembled on a transcript at the time of transcription, thus determining the translational fate of the transcript (41). A highly conserved family of Y-box proteins is found in cytoplasmic messenger ribonucleoprotein particles, where the proteins are thought to play a role in restricting the recruitment of mRNA to the translational machinery (41–43). The major mRNA-associated protein, YB-1, destabilizes the interaction of eIF4E and the 5′ mRNA cap *in vitro*, and overexpression of YB-1 results in translational repression *in vivo* (44). Thus, alterations in RNA packaging can also play an important role in translational regulation.

Translation Alterations Encountered in Cancer

Three main alterations at the translational level occur in cancer: variations in mRNA sequences that increase or decrease translational efficiency, changes in the expression or availability of components of the translational machinery, and activation of translation through aberrantly activated signal transduction pathways. The first alteration affects the translation of an individual mRNA that may play a role in carcinogenesis. The second and third alterations can lead to more global changes, such as an increase in the overall rate of protein synthesis, and the translational activation of several mRNA species.

Variations in mRNA Sequence

Variations in mRNA sequence affect the translational efficiency of the transcript. A brief description of these variations and examples of each mechanism follow.

Mutations. Mutations in the mRNA sequence, especially in the 5′ UTR, can alter its translational efficiency, as seen in the following examples.

c-myc. Saito *et al.* proposed that translation of full-length c-*myc* is repressed, whereas in several Burkitt lymphomas that have deletions of the mRNA 5′ UTR, translation of *c-myc* is more efficient (45). More recently, it was reported that the 5′ UTR of *c-myc* contains an IRES, and thus *c-myc* translation can be initiated by a cap-independent as well as a cap-dependent mechanism (46, 47). In patients with multiple myeloma, a C→T mutation in the c-myc IRES was identified (48) and found to cause an enhanced initiation of translation via internal ribosomal entry (49).

BRCA1. A somatic point mutation (117 G→C) in position −3 with respect to the start codon of the *BRCA1* gene was identified in a highly aggressive sporadic breast cancer (50). Chimeric constructs consisting of the wild-type or mutated *BRCA1* 5′ UTR and a downstream luciferase reporter demonstrated a decrease in the translational efficiency with the 5′ UTR mutation.

Cyclin-dependent Kinase Inhibitor 2A. Some inherited melanoma kindreds have a G→T transversion at base −34 of cyclin-dependent kinase inhibitor-2A, which encodes a cyclin-dependent kinase 4/cyclin-dependent kinase 6 kinase inhibitor important in G_1 checkpoint regulation (51). This mutation gives rise to a novel AUG translation initiation codon, creating an upstream open reading frame that competes for scanning ribosomes and decreases translation from the wild-type AUG.

Alternate Splicing and Alternate Transcription Start Sites. Alterations in splicing and alternate transcription sites can lead to variations in 5′ UTR sequence, length, and secondary structure, ultimately impacting translational efficiency.

ATM. The *ATM* gene has four noncoding exons in its 5′ UTR that undergo extensive alternative splicing (52). The contents of 12 different 5′ UTRs that show considerable diversity in length and sequence have been identified. These divergent 5′ leader sequences play an important role in the translational regulation of the *ATM* gene.

mdm. In a subset of tumors, overexpression of the oncoprotein mdm2 results in enhanced translation of the *mdm2* mRNA. Use of different promoters leads to two *mdm2* transcripts that differ only in their 5′ leaders (53). The longer 5′ UTR contains two upstream open reading frames, and this mRNA is loaded with ribosomes inefficiently compared with the short 5′ UTR.

BRCA1. In a normal mammary gland, *BRCA1* mRNA is expressed with a shorter leader sequence (5′UTRa), whereas in sporadic breast cancer tissue, *BRCA1* mRNA is expressed with a longer leader sequence (5′ UTRb); the translational efficiency of transcripts containing 5′ UTRb is 10 times lower than that of transcripts containing 5′ UTRa (54).

TGF-β3. *TGF-β3* mRNA includes a 1.1-kb 5′ UTR, which exerts an inhibitory effect on translation. Many human breast cancer cell lines contain a novel *TGF-β3* transcript with a 5′ UTR that is 870 nucleotides shorter and has a 7-fold greater translational efficiency than the normal *TGF-β3* mRNA (55).

Alternate Polyadenylation Sites. Multiple polyadenylation signals leading to the generation of several transcripts with differing 3′ UTR have been described for several mRNA species, such as the *RET* proto-oncogene (56), *ATM* gene (52), tissue inhibitor of metalloproteinases-3 (57), *RHOA* proto-oncogene (58), and calmodulin-I (59). Although the effect of these alternate 3′ UTRs on translation is not yet known, they may be important in RNA-protein interactions that affect translational recruitment. The role of these alterations in cancer development and progression is unknown.

Alterations in the Components of the Translation Machinery

Alterations in the components of translation machinery can take many forms.

Overexpresssion of eIF4E. Overexpression of eIF4E causes malignant transformation in rodent cells (60) and the deregulation of HeLa cell growth (61). Polunovsky *et al.* (62) found that eIF4E overexpression substitutes for serum and individual growth factors in preserving viability of fibroblasts, which suggests that eIF4E can mediate both proliferative and survival signaling.

Elevated levels of eIF4E mRNA have been found in a broad spectrum of transformed cell lines (63). eIF4E levels are elevated in all ductal carcinoma *in situ* specimens and invasive ductal carcinomas, compared with benign breast specimens evaluated with Western blot analysis (64, 65). Preliminary studies suggest that this overexpression is attributable to gene amplification (66).

There are accumulating data suggesting that eIF4E overexpression can be valuable as a prognostic marker. eIF4E overexpression was found in a retrospective study to be a marker of poor prognosis in stages I to III breast carcinoma (67). Verification of the prognostic value of eIF4E in breast cancer is now under way in a prospective trial (67). However, in a different study, eIF4E expression was correlated with the aggressive behavior of non-Hodgkin's lymphomas (68). In a prospective analysis of patients with head and neck cancer, elevated levels of eIF4E in histologically tumor-free surgical margins predicted a significantly increased risk of local-regional recurrence (9). These results all suggest that eIF4E overexpression can be used to select patients who might benefit from more aggressive systemic therapy. Furthermore, the head and neck cancer data suggest that eIF4E overexpression is a field defect and can be used to guide local therapy.

Alterations in Other Initiation Factors. Alterations in a number of other initiation factors have been associated with cancer. Overproduction of eIF4G, similar to eIF4E, leads to malignant transformation *in vitro* (69). eIF-2α is found in increased levels in bronchioloalveolar carcinomas of the lung (3). Initiation factor eIF-4A1 is overexpressed in melanoma (70) and hepatocellular carcinoma (71). The p40 subunit of translation initiation factor 3 is amplified and overexpressed in breast and prostate cancer (72), and the eIF3-p110 subunit is overexpressed in testicular seminoma (73). The role that overexpression of these initiation factors plays on the development and progression of cancer, if any, is not known.

Overexpression of S6K. S6K is amplified and highly overexpressed in the MCF7 breast cancer cell line, compared with normal mammary epithelium (74). In a study by Barlund *et al.* (74), S6K was amplified in 59 of 668 primary breast tumors, and a statistically significant association was observed between amplification and poor prognosis.

Overexpression of PAP. PAP catalyzes 3′ poly(A) synthesis. PAP is overexpressed in human cancer cells compared with normal and virally transformed cells (75). PAP enzymatic activity in breast tumors has been correlated with PAP protein levels (76) and, in mammary tumor cytosols, was found to be an independent factor for predicting survival (76). Little is known, however, about how PAP expression or activity affects the translational profile.

Alterations in RNA-binding Proteins. Even less is known about alterations in RNA packaging in cancer. Increased expression and nuclear localization of the RNA-binding protein YB-1 are indicators of a poor prognosis for breast cancer (77), non-small cell lung cancer (78), and ovarian cancer (79). However, this effect may be mediated at least in part at the level of transcription, because YB-1 increases chemoresistance by enhancing the transcription of a multidrug resistance gene (80).

Activation of Signal Transduction Pathways

Activation of signal transduction pathways by loss of tumor suppressor genes or overexpression of certain tyrosine kinases can contribute to the growth and aggressiveness of tumors. An important mutant in human cancers is the tumor suppressor gene *PTEN*, which leads to the activation of the PI3K/Akt pathway. Activation of PI3K and Akt induces the oncogenic transformation of chicken embryo fibroblasts. The transformed cells show constitutive phosphorylation of S6K and of 4E-BP1 (81). A mutant Akt that retains kinase activity but does not phosphorylate S6K or 4E-BP1 does not transform fibroblasts, which suggests a correlation between the oncogenicity of PI3K and Akt and the phosphorylation of S6K and 4E-BP1 (81).

Several tyrosine kinases such as platelet-derived growth factor, insulin-like growth factor, HER2/neu, and epidermal growth factor receptor are overexpressed in cancer. Because these kinases activate downstream signal transduction pathways known to alter translation initiation, activation of translation is likely to contribute to the growth and aggressiveness of these tumors. Furthermore, the mRNA for many of these kinases themselves are under translational control. For example, HER2/neu mRNA is translationally controlled both by a short upstream open reading frame that represses HER2/neu translation in a cell type-independent manner and by a distinct cell type-dependent mechanism that increases translational efficiency (82). HER2/neu translation is different in transformed and normal cells. Thus, it is possible that alterations at the translational level can in part account for the discrepancy between *HER2/neu* gene amplification detected by fluorescence *in situ* hybridization and protein levels detected by immunohistochemical assays.

Translation Targets of Selected Cancer Therapy

Components of the translation machinery and signal pathways involved in the activation of translation initiation represent good targets for cancer therapy.

Targeting the mTOR Signaling Pathway: Rapamycin and Tumstatin

Rapamycin inhibits the proliferation of lymphocytes. It was initially developed as an immunosuppressive drug for organ transplantation. Rapamycin with FKBP 12 (FK506-binding protein, M_r 12,000) binds to mTOR to inhibit its function.

Rapamycin causes a small but significant reduction in the initiation rate of protein synthesis (83). It blocks cell growth in part by blocking S6 phosphorylation and selectively suppressing the translation of 5′ TOP mRNAs, such as ribosomal proteins, and elongation factors (83–85). Rapamycin also blocks 4E-BP1 phosphorylation and inhibits cap-dependent but not cap-independent translation (17, 86).

The rapamycin-sensitive signal transduction pathway, activated during malignant transformation and cancer progression, is now being studied as a target for cancer therapy (87). Prostate, breast, small cell lung, glioblastoma, melanoma, and T-cell leukemia are among the cancer lines most sensitive to the rapamycin analogue CCI-779 (Wyeth-Ayerst Research; Ref. 87). In rhabdomyosarcoma cell lines, rapamycin is either cytostatic or cytocidal, depending on the p53 status of the cell; p53 wild-type cells treated with rapamycin arrest in the G_1 phase and maintain their viability, whereas p53 mutant cells accumulate in G_1 and undergo apoptosis (88, 89). In a recently reported study using human primitive neuroectodermal tumor and medulloblastoma models, rapamycin exhibited more cytotoxicity in combination with cisplatin and camptothecin than as a single agent. *In vivo*, CCI-779 delayed growth of xenografts by 160% after 1 week of therapy and 240% after 2 weeks. A single high-dose administration caused a 37% decrease in tumor volume. Growth inhibition *in vivo* was 1.3 times greater, with cisplatin in combination with CCI-779 than with cisplatin alone (90). Thus, preclinical studies suggest that rapamycin analogues are useful as single agents and in combination with chemotherapy.

Rapamycin analogues CCI-779 and RAD001 (Novartis, Basel, Switzerland) are now in clinical trials. Because of the known effect of rapamycin on lymphocyte proliferation, a potential problem with rapamycin analogues is immunosuppression. However, although prolonged immunosuppression can result from rapamycin and CCI-779 administered on continuous-dose schedules, the immunosuppressive effects of rapamycin analogues resolve in ~24 h after therapy (91). The principal toxicities of CCI-779 have included dermatological toxicity, myelosuppression, infection, mucositis, diarrhea, reversible elevations in liver function tests, hyperglycemia, hypokalemia, hypocalcemia, and depression (87, 92–94). Phase II trials of CCI-779 have been conducted in advanced renal cell carcinoma and in stage III/IV breast carcinoma patients who failed with prior chemotherapy. In the results reported in abstract form, although there were no complete responses, partial responses were documented in both renal cell carcinoma and in breast carcinoma (94, 95). Thus, CCI-779 has documented preliminary clinical activity in a previously treated, unselected patient population.

Active investigation is under way into patient selection for mTOR inhibitors. Several studies have found an enhanced efficacy of CCI-779 in PTEN-null tumors (30, 96). Another study found that six of eight breast cancer cell lines were responsive to CCI-779, although only two of these lines lacked PTEN (97) There was, however, a positive correlation between Akt activation and CCI-779 sensitivity (97). This correlation suggests that activation of the PI3K-Akt pathway,

regardless of whether it is attributable to a PTEN mutation or to overexpression of receptor tyrosine kinases, makes cancer cell amenable to mTOR-directed therapy. In contrast, lower levels of the target of mTOR, 4E-BP1, are associated with rapamycin resistance; thus, a lower 4E-BP1/eIF4E ratio may predict rapamycin resistance (98).

Another mode of activity for rapamycin and its analogues appears to be through inhibition of angiogenesis. This activity may be both through direct inhibition of endothelial cell proliferation as a result of mTOR inhibition in these cells or by inhibition of translation of such proangiogenic factors as vascular endothelial growth factor in tumor cells (99, 100).

The angiogenesis inhibitor tumstatin, another anticancer drug currently under study, was also found recently to inhibit translation in endothelial cells (101). Through a requisite interaction with integrin, tumstatin inhibits activation of the PI3K/Akt pathway and mTOR in endothelial cells and prevents dissociation of eIF4E from 4E-BP1, thereby inhibiting cap-dependent translation. These findings suggest that endothelial cells are especially sensitive to therapies targeting the mTOR-signaling pathway.

Targeting eIF2α: EPA, Clotrimazole, mda-7, and Flavonoids

EPA is an n-3 polyunsaturated fatty acid found in the fish-based diets of populations having a low incidence of cancer (102). EPA inhibits the proliferation of cancer cells (103), as well as in animal models (104, 105). It blocks cell division by inhibiting translation initiation (105). EPA releases Ca^{2+} from intracellular stores while inhibiting their refilling, thereby activating PKR. PKR, in turn phosphorylates and inhibits eIF2α, resulting in the inhibition of protein synthesis at the level of translation initiation. Similarly, clotrimazole, a potent antiproliferative agent *in vitro* and *in vivo*, inhibits cell growth through depletion of Ca^{2+} stores, activation of PKR, and phosphorylation of eIF2α (106). Consequently, clotrimazole preferentially decreases the expression of cyclins A, E, and D1, resulting in blockage of the cell cycle in G_1.

mda-7 is a novel tumor suppressor gene being developed as a gene therapy agent. Adenoviral transfer of *mda-7* (Ad-mda7) induces apoptosis in many cancer cells including breast, colorectal, and lung cancer (107–109). Ad-mda7 also induces and activates PKR, which leads to phosphorylation of eIF2α and induction of apoptosis (110).

Flavonoids such as genistein and quercetin suppress tumor cell growth. All three mammalian eIF2α kinases, PKR, heme-regulated inhibitor, and PERK/PEK, are activated by flavonoids, with phosphorylation of eIF2α and inhibition of protein synthesis (111).

Targeting eIF4A and eIF4E: Antisense RNA and Peptides

Antisense expression of eIF4A decreases the proliferation rate of melanoma cells (112). Sequestration of eIF4E by overexpression of 4E-BP1 is proapoptotic and decreases tumorigenicity (113, 114). Reduction of eIF4E with antisense RNA decreases soft agar growth, increases tumor latency, and increases the rates of tumor doubling times (7). Antisense eIF4E RNA treatment also reduces the expression of angiogenic factors (115) and has been proposed as a potential adjuvant therapy for head and neck cancers, particularly when elevated eIF4E is found in surgical margins. Small molecule inhibitors that bind the eIF4G/4E-BP1-binding domain of eIF4E are proapoptotic (116) and are also being actively pursued.

Exploiting Selective Translation for Gene Therapy

A different therapeutic approach that takes advantage of the enhanced cap-dependent translation in cancer cells is the use of gene therapy vectors encoding suicide genes with highly structured 5′ UTR. These mRNA would thus be at a competitive disadvantage in normal cells and not translate well, whereas in cancer cells, they would translate more efficiently. For example, the introduction of the 5′ UTR of fibroblast growth factor-2 5′ to the coding sequence of *herpes simplex virus type-1 thymidine kinase* gene, allows for selective translation of *herpes simplex virus type-1 thymidine kinase* gene in breast cancer cell lines compared with normal mammary cell lines and results in selective sensitivity to ganciclovir (117).

Toward the Future

Translation is a crucial process in every cell. However, several alterations in translational control occur in cancer. Cancer cells appear to need an aberrantly activated translational state for survival, thus allowing the targeting of translation initiation with surprisingly low toxicity. Components of the translational machinery, such as eIF4E, and signal transduction pathways involved in translation initiation, such mTOR, represent promising targets for cancer therapy. Inhibitors of the mTOR have already shown some preliminary activity in clinical trials. It is possible that with the development of better predictive markers and better patient selection, response rates to single-agent therapy can be improved. Similar to other cytostatic agents, however, mTOR inhibitors are most likely to achieve clinical utility in combination therapy. In the interim, our increasing understanding of translation initiation and signal transduction pathways promise to lead to the identification of new therapeutic targets in the near future.

Acknowledgments

We thank Gayle Nesom from The University of Texas M. D. Anderson Cancer Center Department of Scientific Publications for editorial assistance and Dr. Elmer Bernstam for assistance with manuscript preparation.

References

1. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. Molecular mechanisms of translation initiation in eukaryotes. Proc. Natl. Acad. Sci. USA, *98*: 7029–7036, 2001.

2. Rosenwald, I. B., Kaspar, R., Rousseau, D., Gehrke, L., Leboulch, P., Chen, J. J., Schmidt, E. V., Sonenberg, N., and London, I. M. Eukaryotic translation initiation factor 4E regulates expression of cyclin D1 at transcriptional and post-transcriptional levels. J. Biol. Chem., *270*: 21176–21180, 1995.

3. Rosenwald, I. B., Hutzler, M. J., Wang, S., Savas, L., and Fraire, A. E. Expression of eukaryotic translation initiation factors 4E and 2α is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. Cancer (Phila.), *92*: 2164–2171, 2001.

4. Darveau, A., Pelletier, J., and Sonenberg, N. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5′ untranslated region. Proc. Natl. Acad. Sci. USA, *82:* 2315–2319, 1985.

5. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA, *83:* 2850–2854, 1986.

6. Koromilas, A. E., Lazaris-Karatzas, A., and Sonenberg, N. mRNAs containing extensive secondary structure in their 5′ non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. EMBO J., *11:* 4153–4158, 1992.

7. Rinker-Schaeffer, C. W., Graff, J. R., De Benedetti, A., Zimmer, S. G., and Rhoads, R. E. Decreasing the level of translation initiation factor 4E with antisense RNA causes reversal of ras-mediated transformation and tumorigenesis of cloned rat embryo fibroblasts. Int. J. Cancer, *55:* 841–847, 1993.

8. Kevil, C. G., De Benedetti, A., Payne, D. K., Coe, L. L., Laroux, F. S., and Alexander, J. S. Translational regulation of vascular permeability factor by eukaryotic initiation factor 4E: implications for tumor angiogenesis. Int. J. Cancer, *65:* 785–790, 1996.

9. Nathan, C. A., Franklin, S., Abreo, F. W., Nassar, R., De Benedetti, A., and Glass, J. Analysis of surgical margins with the molecular marker eIF4E: a prognostic factor in patients with head and neck cancer. J. Clin. Oncol., *17:* 2909–2914, 1999.

10. Fukunaga, R., and Hunter, T. MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. EMBO J., *16:* 1921–1933, 1997.

11. Waskiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. EMBO J., *16:* 1909–1920, 1997.

12. Wang, X., Flynn, A., Waskiewicz, A. J., Webb, B. L., Vries, R. G., Baines, I. A., Cooper, J. A., and Proud, C. G. The phosphorylation of eukaryotic initiation factor eIF4E in response to phorbol esters, cell stresses, and cytokines is mediated by distinct MAP kinase pathways. J. Biol. Chem., *273:* 9373–9377, 1998.

13. Pyronnet, S., Imataka, H., Gingras, A. C., Fukunaga, R., Hunter, T., and Sonenberg, N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. EMBO J., *18:* 270–279, 1999.

14. Klein, M., Scheper, G. C., Voorma, H. O., and Thomas, A. A. Regulation of translation initiation factors by signal transduction. Eur. J. Biochem., *253:* 531–544, 1998.

15. Raught, B., and Gingras, A. C. eIF4E activity is regulated at multiple levels. Int. J. Biochem. Cell Biol., *31:* 43–57, 1999.

16. Takeuchi, K., Shibamoto, S., Nagamine, K., Shigemori, I., Omura, S., Kitamura, N., and Ito, F. Signaling pathways leading to transcription and translation cooperatively regulate the transient increase in expression of c-Fos protein. J. Biol. Chem., *276:* 26077–26083, 2001.

17. Kawasome, H., Papst, P., Webb, S., Keller, G. M., Johnson, G. L., Gelfand, E. W., and Terada, N. Targeted disruption of p70(s6k) defines its role in protein synthesis and rapamycin sensitivity. Proc. Natl. Acad. Sci. USA, *95:* 5033–5038, 1998.

18. Christie, G. R., Hajduch, E., Hundal, H. S., Proud, C. G., and Taylor, P. M. Intracellular sensing of amino acids in *Xenopus laevis* oocytes stimulates p70 S6 kinase in a target of rapamycin-dependent manner. J. Biol. Chem., *277:* 9952–9957, 2002.

19. Hara, K., Yonezawa, K., Weng, Q. P., Kozlowski, M. T., Belham, C., and Avruch, J. Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. J. Biol. Chem., *273:* 14484–14494, 1998.

20. Graves, L. M., Bornfeldt, K. E., Argast, G. M., Krebs, E. G., Kong, X., Lin, T. A., and Lawrence, J. C., Jr. cAMP- and rapamycin-sensitive regulation of the association of eukaryotic initiation factor 4E and the translational regulator PHAS-I in aortic smooth muscle cells. Proc. Natl. Acad. Sci. USA, *92:* 7222–7226, 1995.

21. Merrick, W. C., and Hershey, J. W. B. The pathway and mechanism of eukaryotic protein synthesis. *In:* J. W. B. Hershey and M. B. Mathews (eds.), Translational Control, pp. 31–69. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1996.

22. Kimball, S. R. Eukaryotic initiation factor eIF2. Int. J. Biochem. Cell Biol., *31:* 25–29, 1999.

23. Jagus, R., Joshi, B., and Barber, G. N. PKR, apoptosis and cancer. Int. J. Biochem. Cell Biol., *31:* 123–138, 1999.

24. Thomas, G., and Hall, M. N. TOR signalling and control of cell growth. Curr. Opin. Cell Biol., *9:* 782–787, 1997.

25. Gingras, A. C., Raught, B., and Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. Genes Dev., *15:* 807–826, 2001.

26. Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R., and Sonenberg, N. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. Genes Dev., *13:* 1422–1437, 1999.

27. Kumar, V., Pandey, P., Sabatini, D., Kumar, M., Majumder, P. K., Bharti, A., Carmichael, G., Kufe, D., and Kharbanda, S. Functional interaction between RAFT1/FRAP/mTOR and protein kinase Cδ in the regulation of cap-dependent initiation of translation. EMBO J., *19:* 1087–1097, 2000.

28. Yang, D. Q., and Kastan, M. B. Participation of ATM in insulin signalling through phosphorylation of eIF-4E-binding protein 1. Nat. Cell. Biol., *2:* 893–898, 2000.

29. Liu, G., Zhang, Y., Bode, A. M., Ma, W. Y., and Dong, Z. Phosphorylation of 4E-BP1 is mediated by the p38/MSK1 pathway in response to UVB irradiation. J. Biol. Chem., *277:* 8810–8816, 2002.

30. Neshat, M. S., Mellinghoff, I. K., Tran, C., Stiles, B., Thomas, G., Petersen, R., Frost, P., Gibbons, J. J., Wu, H., and Sawyers, C. L. Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. Proc. Natl. Acad. Sci. USA, *98:* 10314–10319, 2001.

31. Sekulic, A., Hudson, C. C., Homme, J. L., Yin, P., Otterness, D. M., Karnitz, L. M., and Abraham, R. T. A direct linkage between the phosphoinositide 3-kinase-AKT signaling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. Cancer Res., *60:* 3504–3513, 2000.

32. Scott, P. H., and Lawrence, J. C., Jr. Attenuation of mammalian target of rapamycin activity by increased cAMP in 3T3-L1 adipocytes. J. Biol. Chem., *273:* 34498–34501, 1998.

33. Reynolds, T. T., Bodine, S. C., and Lawrence, J. C., Jr. Control of Ser2448 phosphorylation in the mammalian target of rapamycin by insulin and skeletal muscle load. J. Biol. Chem., *277:* 17657–17662, 2002.

34. Peterson, R. T., Beal, P. A., Comb, M. J., and Schreiber, S. L. FKBP12-rapamycin-associated protein (FRAP) autophosphorylates at serine 2481 under translationally repressive conditions. J. Biol. Chem., *275:* 7416–7423, 2000.

35. Peterson, R. T., Desai, B. N., Hardwick, J. S., and Schreiber, S. L. Protein phosphatase 2A interacts with the 70-kDa S6 kinase and is activated by inhibition of FKBP12-rapamycin-associated protein. Proc. Natl. Acad. Sci. USA, *96:* 4438–4442, 1999.

36. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. Genes Dev., *3:* 803–815, 1989.

37. Sheets, M. D., Wu, M., and Wickens, M. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. Nature (Lond.), *374:* 511–516, 1995.

38. Varnum, S. M., and Wormington, W. M. Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific *cis*-sequences: a default mechanism for translational control. Genes Dev., *4:* 2278–2286, 1990.

39. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev., *5:* 2108–2116, 1991.

40. Sachs, A. B., and Varani, G. Eukaryotic translation initiation: there are (at least) two sides to every story. Nat. Struct. Biol., *7:* 356–361, 2000.

41. Wolffe, A. P., and Meric, F. Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. Int. J. Biochem. Cell Biol., *28:* 247–257, 1996.

42. Evdokimova, V. M., Wei, C. L., Sitikov, A. S., Simonenko, P. N., Lazarev, O. A., Vasilenko, K. S., Ustinov, V. A., Hershey, J. W., and Ovchinnikov, L. P. The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. J. Biol. Chem., *270:* 3186–3192, 1995.

43. Matsumoto, K., Meric, F., and Wolffe, A. P. Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. J. Biol. Chem., *271:* 22706–22712, 1996.

44. Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svitkin, Y., Ovchinnikov, L. P., and Sonenberg, N. The major mRNA-associated protein YB-1 is a potent 5′ cap-dependent mRNA stabilizer. EMBO J., *20:* 5491–5502, 2001.

45. Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Tonegawa, S. Activation of the c-*myc* gene by translocation: a model for translational control. Proc. Natl. Acad. Sci. USA, *80:* 7476–7480, 1983.

46. Nanbru, C., Lafon, I., Audigier, S., Gensac, M. C., Vagner, S., Huez, G., and Prats, A. C. Alternative translation of the proto-oncogene c-*myc* by an internal ribosome entry site. J. Biol. Chem., *272:* 32061–32066, 1997.

47. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A., and Willis, A. E. c-Myc 5′ untranslated region contains an internal ribosome entry segment. Oncogene, *16:* 423–428, 1998.

48. Paulin, F. E., West, M. J., Sullivan, N. F., Whitney, R. L., Lyne, L., and Willis, A. E. Aberrant translational control of the c-*myc* gene in multiple myeloma. Oncogene, *13:* 505–513, 1996.

49. Chappell, S. A., LeQuesne, J. P., Paulin, F. E., de Schoolmeester, M. L., Stoneley, M., Soutar, R. L., Ralston, S. H., Helfrich, M. H., and Willis, A. E. A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. Oncogene, *19:* 4437–4440, 2000.

50. Signori, E., Bagni, C., Papa, S., Primerano, B., Rinaldi, M., Amaldi, F., and Fazio, V. M. A somatic mutation in the 5′UTR of *BRCA1* gene in sporadic breast cancer causes down-modulation of translation efficiency. Oncogene, *20:* 4596–4600, 2001.

51. Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. Mutation of the CDKN2A 5′ UTR creates an aberrant initiation codon and predisposes to melanoma. Nat. Genet., *21:* 128–132, 1999.

52. Savitsky, K., Platzer, M., Uziel, T., Gilad, S., Sartiel, A., Rosenthal, A., Elroy-Stein, O., Shiloh, Y., and Rotman, G. Ataxia-telangiectasia: structural diversity of untranslated sequences suggests complex post-transcriptional regulation of *ATM* gene expression. Nucleic Acids Res., *25:* 1678–1684, 1997.

53. Brown, C. Y., Mize, G. J., Pineda, M., George, D. L., and Morris, D. R. Role of two upstream open reading frames in the translational control of oncogene *mdm2*. Oncogene, *18:* 5631–5637, 1999.

54. Sobczak, K., and Krzyzosiak, W. J. Structural determinants of BRCA1 translational regulation. J. Biol. Chem., *277:* 17349–17358, 2002.

55. Arrick, B. A., Grendell, R. L., and Griffin, L. A. Enhanced translational efficiency of a novel transforming growth factor β3 mRNA in human breast cancer cells. Mol. Cell. Biol., *14:* 619–628, 1994.

56. Myers, S. M., Eng, C., Ponder, B. A., and Mulligan, L. M. Characterization of *RET* proto-oncogene 3′ splicing variants and polyadenylation sites: a novel C-terminus for RET. Oncogene, *11:* 2039–2045, 1995.

57. Byrne, J. A., Tomasetto, C., Rouyer, N., Bellocq, J. P., Rio, M. C., and Basset, P. The tissue inhibitor of *metalloproteinases-3* gene in breast carcinoma: identification of multiple polyadenylation sites and a stromal pattern of expression. Mol. Med., *1:* 418–427, 1995.

58. Moscow, J. A., He, R., Gudas, J. M., and Cowan, K. H. Utilization of multiple polyadenylation signals in the human *RHOA* protooncogene. Gene (Amst.), *144:* 229–236, 1994.

59. Senterre-Lesenfants, S., Alag, A. S., and Sobel, M. E. Multiple mRNA species are generated by alternate polyadenylation from the human *calmodulin-1* gene. J. Cell. Biochem., *58:* 445–454, 1995.

60. Lazaris-Karatzas, A., Montine, K. S., and Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5′ cap. Nature (Lond.), *345:* 544–547, 1990.

61. De Benedetti, A., and Rhoads, R. E. Overexpression of eukaryotic protein synthesis initiation factor 4E in HeLa cells results in aberrant growth and morphology. Proc. Natl. Acad. Sci. USA, *87:* 8212–8216, 1990.

62. Polunovsky, V. A., Rosenwald, I. B., Tan, A. T., White, J., Chiang, L., Sonenberg, N., and Bitterman, P. B. Translational control of programmed cell death: eukaryotic translation initiation factor 4E blocks apoptosis in growth-factor-restricted fibroblasts with physiologically expressed or deregulated Myc. Mol. Cell. Biol., *16:* 6573–6581, 1996.

63. Miyagi, Y., Sugiyama, A., Asai, A., Okazaki, T., Kuchino, Y., and Kerr, S. J. Elevated levels of eukaryotic translation initiation factor eIF-4E, mRNA in a broad spectrum of transformed cell lines. Cancer Lett., *91:* 247–252, 1995.

64. Kerekatte, V., Smiley, K., Hu, B., Smith, A., Gelder, F., and De Benedetti, A. The proto-oncogene/translation factor *eIF4E*: a survey of its expression in breast carcinomas. Int. J. Cancer, *64:* 27–31, 1995.

65. Li, B. D., Liu, L., Dawson, M., and De Benedetti, A. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. Cancer (Phila.), *79:* 2385–2390, 1997.

66. Sorrells, D. L., Black, D. R., Meschonat, C., Rhoads, R., De Benedetti, A., Gao, M., Williams, B. J., and Li, B. D. Detection of *eIF4E* gene amplification in breast cancer by competitive PCR. Ann. Surg. Oncol., *5:* 232–237, 1998.

67. Li, B. D., McDonald, J. C., Nassar, R., and De Benedetti, A. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. Ann. Surg., *227:* 756–761; discussion, 761–763, 1998.

68. Wang, S., Rosenwald, I. B., Hutzler, M. J., Pihan, G. A., Savas, L., Chen, J. J., and Woda, B. A. Expression of the eukaryotic translation initiation factors 4E and 2α in non-Hodgkin's lymphomas. Am. J. Pathol., *155:* 247–255, 1999.

69. Fukuchi-Shimogori, T., Ishii, I., Kashiwagi, K., Mashiba, H., Ekimoto, H., and Igarashi, K. Malignant transformation by overproduction of translation initiation factor eIF4G. Cancer Res., *57:* 5041–5044, 1997.

70. Eberle, J., Krasagakis, K., and Orfanos, C. E. Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells *in vitro*. Int. J. Cancer, *71:* 396–401, 1997.

71. Shuda, M., Kondoh, N., Tanaka, K., Ryo, A., Wakatsuki, T., Hada, A., Goseki, N., Igari, T., Hatsuse, K., Aihara, T., Horiuchi, S., Shichita, M., Yamamoto, N., and Yamamoto, M. Enhanced expression of translation factor mRNAs in hepatocellular carcinoma. Anticancer Res., *20:* 2489–2494, 2000.

72. Nupponen, N. N., Porkka, K., Kakkola, L., Tanner, M., Persson, K., Borg, A., Isola, J., and Visakorpi, T. Amplification and overexpression of p40 subunit of eukaryotic translation initiation factor 3 in breast and prostate cancer. Am. J. Pathol., *154:* 1777–1783, 1999.

73. Rothe, M., Ko, Y., Albers, P., and Wernert, N. Eukaryotic initiation factor 3 p110 mRNA is overexpressed in testicular seminomas. Am. J. Pathol., *157:* 1597–1604, 2000.

74. Barlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., Kallioniemi, O. P., and Kallioniemi, A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J. Natl. Cancer Inst. (Bethesda), *92:* 1252–1259, 2000.

75. Topalian, S. L., Kaneko, S., Gonzales, M. I., Bond, G. L., Ward, Y., and Manley, J. L. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. Mol. Cell. Biol., *21:* 5614–5623, 2001.

76. Scorilas, A., Talieri, M., Ardavanis, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsiapalis, C. M., and Trangas, T. Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: a new independent prognostic marker in primary breast cancer. Cancer Res., *60:* 5427–5433, 2000.

77. Janz, M., Harbeck, N., Dettmar, P., Berger, U., Schmidt, A., Jurchott, K., Schmitt, M., and Royer, H. D. Y-box factor YB-1 predicts drug resistance and patient outcome in breast cancer independent of clinically relevant tumor biologic factors HER2, uPA and PAI-1. Int. J. Cancer, *97:* 278–282, 2002.

78. Shibahara, K., Sugio, K., Osaki, T., Uchiumi, T., Maehara, Y., Kohno, K., Yasumoto, K., Sugimachi, K., and Kuwano, M. Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer. Clin. Cancer Res., *7:* 3151–3155, 2001.

79. Kamura, T., Yahata, H., Amada, S., Ogawa, S., Sonoda, T., Kobayashi, H., Mitsumoto, M., Kohno, K., Kuwano, M., and Nakano, H. Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma? Cancer (Phila.), *85:* 2450–2454, 1999.

80. Bargou, R. C., Jurchott, K., Wagener, C., Bergmann, S., Metzner, S., Bommert, K., Mapara, M. Y., Winzer, K. J., Dietel, M., Dorken, B., and Royer, H. D. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic *MDR1* gene expression. Nat. Med., *3:* 447–450, 1997.

81. Aoki, M., Blazek, E., and Vogt, P. K. A role of the kinase mTOR in cellular transformation induced by the oncoproteins P3k and Akt. Proc. Natl. Acad. Sci. USA, *98:* 136–141, 2001.

82. Child, S. J., Miller, M. K., and Geballe, A. P. Cell type-dependent and -independent control of HER-2/neu translation. Int. J. Biochem. Cell Biol., *31:* 201–213, 1999.

83. Jefferies, H. B., Reinhard, C., Kozma, S. C., and Thomas, G. Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family. Proc. Natl. Acad. Sci. USA, *91:* 4441–4445, 1994.

84. Terada, N., Patel, H. R., Takase, K., Kohno, K., Nairn, A. C., and Gelfand, E. W. Rapamycin selectively inhibits translation of mRNAs encoding elongation factors and ribosomal proteins. Proc. Natl. Acad. Sci. USA, *91:* 11477–11481, 1994.

85. Jefferies, H. B., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. EMBO J., *16:* 3693–3704, 1997.

86. Beretta, L., Gingras, A. C., Svitkin, Y. V., Hall, M. N., and Sonenberg, N. Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. EMBO J., *15:* 658–664, 1996.

87. Hidalgo, M., and Rowinsky, E. K. The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. Oncogene, *19:* 6680–6686, 2000.

88. Hosoi, H., Dilling, M. B., Shikata, T., Liu, L. N., Shu, L., Ashmun, R. A., Germain, G. S., Abraham, R. T., and Houghton, P. J. Rapamycin causes poorly reversible inhibition of mTOR and induces p53-independent apoptosis in human rhabdomyosarcoma cells. Cancer Res., *59:* 886–894, 1999.

89. Huang, S., and Houghton, P. J. Resistance to rapamycin: a novel anticancer drug. Cancer Metastasis Rev., *20:* 69–78, 2001.

90. Geoerger, B., Kerr, K., Tang, C. B., Fung, K. M., Powell, B., Sutton, L. N., Phillips, P. C., and Janss, A. J. Antitumor activity of the rapamycin analog CCI-779 in human primitive neuroectodermal tumor/medulloblastoma models as single agent and in combination chemotherapy. Cancer Res., *61:* 1527–1532, 2001.

91. Gibbons, J. J., Discafani, C., Peterson, R., Hernandez, R., Skotnicki, J., and Frost, P. The effect of CCI-779, a novel macrolide anti-tumor agent, on the growth of human tumor cells *in vitro* and in nude mouse xenografts *in vivo*. Proc. Am. Assoc. Cancer Res., *40:* 301, 1999.

92. Hidalgo, M., Rowinsky, E., Erlichman, C., Marshall, B., Marks, R., Edwards, T., and Buckner, J. J. A Phase I and pharmacological study of CCI-779 cycle inhibitor. Ann. Oncol., *11* (Suppl. 4): 133, 2001.

93. Alexandre, J., Raymond, E., Depenbrock, H., Makhaldi, S., Angevin, E., Paillet, C., Hanauske, A., Frisch, J., Feussner, A., and Armand, J. P. CCI-779, a new rapamycin analog, has antitumor activity at doses inducing only mild cutaneous effects and mucositis: early results of an ongoing Phase I study. Proceedings of the 1999 AACR-NCI-EORTC International Conference, Clin. Cancer Res., *5* (Suppl.): 3730s, 1999.

94. Chan, S., Johnston, S., Scheulen, M. E., Mross, K., Morant, A., Lahr, A., Feussner, A., Berger, M., and Kirsch, T. First report: a Phase 2 study of the safety and activity of CCI-779 for patients with locally advanced or metastatic breast cancer failing prior chemotherapy. Proc. Am. Soc. Clin. Oncol., *21:* 44a, 2002.

95. Atkins, M. B., Hidalgo, M., Stadler, W., Logan, T., Dutcher, J. P., Hudes, G., Park, Y., Marshall, B., Boni, J., and Dukart, G. A randomized double-blind Phase 2 study of intravenous CCI-779 administered weekly to patients with advanced renal cell carcinoma. Proc. Am. Soc. Clin. Oncol., *21:* 10a, 2002.

96. Smith, S. G., Trinh, C. M., Inge, L. J., Thomas, G., Cloughsey, T. F., Sawyers, C. L., and Mischel, P. S. PTEN expression status predicts glioblastoma cell sensitivity to CCI-779. Proc. Am. Assoc. Cancer Res., *43:* 335, 2002.

97. Yu, K., Toral-Barza, L., Discafani, C., Zhang, W. G., Skotnicki, J., Frost, P., and Gibbons, J. J. mTOR, a novel target in breast cancer: the effect of CCI-779, an mTOR inhibitor, in preclinical models of breast cancer. Endocr. Relat. Cancer, *8:* 249–258, 2001.

98. Dilling, M. B., Germain, G. S., Dudkin, L., Jayaraman, A. L., Zhang, X., Harwood, F. C., and Houghton, P. J. 4E-binding proteins, the suppressors of eukaryotic initiation factor 4E, are downregulated in cells with acquired or intrinsic resistance to rapamycin. J. Biol. Chem., *277:* 13907–13917, 2002.

99. Guba, M., von Breitenbuch, P., Steinbauer, M., Koehl, G., Flegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., and Geissler, E. K. Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: involvement of vascular endothelial growth factor. Nat. Med., *8:* 128–135, 2002.

100. Lane, H. A., Schell, C., Theuer, A., O'Reilly, T., and Wood, J. Anti-angiogenic activity of RAD001, an orally active anticancer agent. Proc. Am. Assoc. Cancer Res., *43:* 184, 2002.

101. Maeshima, Y., Sudhakar, A., Lively, J. C., Ueki, K., Kharbanda, S., Kahn, C. R., Sonenberg, N., Hynes, R. O., and Kalluri, R. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. Science (Wash. DC), *295:* 140–143, 2002.

102. Caygill, C. P., Charlett, A., and Hill, M. J. Fat, fish, fish oil and cancer. Br. J. Cancer, *74:* 159–164, 1996.

103. Falconer, J. S., Ross, J. A., Fearon, K. C., Hawkins, R. A., O'Riordain, M. G., and Carter, D. C. Effect of eicosapentaenoic acid and other fatty acids on the growth *in vitro* of human pancreatic cancer cell lines. Br. J. Cancer, *69:* 826–832, 1994.

104. Noguchi, M., Minami, M., Yagasaki, R., Kinoshita, K., Earashi, M., Kitagawa, H., Taniya, T., and Miyazaki, I. Chemoprevention of DMBA-induced mammary carcinogenesis in rats by low-dose EPA and DHA. Br. J. Cancer, *75:* 348–353, 1997.

105. Palakurthi, S. S., Fluckiger, R., Aktas, H., Changolkar, A. K., Shahsafaei, A., Harnett, S., Kilic, E., and Halperin, J. A. Inhibition of translation initiation mediates the anticancer effect of the n-3 polyunsaturated fatty acid eicosapentaenoic acid. Cancer Res., *60:* 2919–2925, 2000.

106. Aktas, H., Fluckiger, R., Acosta, J. A., Savage, J. M., Palakurthi, S. S., and Halperin, J. A. Depletion of intracellular Ca^{2+} stores, phosphorylation of eIF2α, and sustained inhibition of translation initiation mediate the anticancer effects of clotrimazole. Proc. Natl. Acad. Sci. USA, *95:* 8280–8285, 1998.

107. Mhashilkar, A. M., Schrock, R. D., Hindi, M., Liao, J., Sieger, K., Kourouma, F., Zou-Yang, X. H., Onishi, E., Takh, O., Vedvick, T. S., Fanger, G., Stewart, L., Watson, G. J., Snary, D., Fisher, P. B., Saeki, T., Roth, J. A., Ramesh, R., and Chada, S. Melanoma differentiation associated gene-7 (*mda-7*): a novel anti-tumor gene for cancer gene therapy. Mol. Med., *7:* 271–282, 2001.

108. Su, Z. Z., Madireddi, M. T., Lin, J. J., Young, C. S., Kitada, S., Reed, J. C., Goldstein, N. I., and Fisher, P. B. The cancer growth suppressor gene *mda-7* selectively induces apoptosis in human breast cancer cells and inhibits tumor growth in nude mice. Proc. Natl. Acad. Sci. USA, *95:* 14400–14405, 1998.

109. Saeki, T., Mhashilkar, A., Chada, S., Branch, C., Roth, J. A., and Ramesh, R. Tumor-suppressive effects by adenovirus-mediated *mda-7* gene transfer in non-small cell lung cancer cell *in vitro*. Gene Ther., *7:* 2051–2057, 2000.

110. Pataer, A., Vorburger, S. A., Barber, G. N., Chada, S., Mhashilkar, A. M., Zou-Yang, H., Stewart, A. L., Balachandran, S., Roth, J. A., Hunt, K. K., and Swisher, S. G. Adenoviral transfer of the melanoma differentiation-associated gene 7 (*mda7*) induces apoptosis of lung cancer cells via up-regulation of the double-stranded RNA-dependent protein kinase (PKR). Cancer Res., *62:* 2239–2243, 2002.

111. Ito, T., Warnken, S. P., and May, W. S. Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2α kinases. Biochem. Biophys. Res. Commun., *265:* 589–594, 1999.

112. Eberle, J., Fecker, L. F., Bittner, J. U., Orfanos, C. E., and Geilen, C. C. Decreased proliferation of human melanoma cell lines caused by antisense RNA against translation factor eIF-4A1. Br. J. Cancer, *86:* 1957–1962, 2002.

113. Polunovsky, V. A., Gingras, A. C., Sonenberg, N., Peterson, M., Tan, A., Rubins, J. B., Manivel, J. C., and Bitterman, P. B. Translational control of the antiapoptotic function of Ras. J. Biol. Chem., *275:* 24776–24780, 2000.

114. D'Cunha, J., Kratzke, M. G., Alter, M. D., Polunovsky, V. A., Bitterman, P. B., and Kratzke, R. A. Over-expression of the translational repressor 4E-BP1 inhibits NSCLC tumorigenicity *in vivo*. Proc. Am. Assoc. Cancer Res., *43:* 816–817, 2002.

115. DeFatta, R. J., Nathan, C. A., and De Benedetti, A. Antisense RNA to eIF4E suppresses oncogenic properties of a head and neck squamous cell carcinoma cell line. Laryngoscope, *110:* 928–933, 2000.

116. Herbert, T. P., Fahraeus, R., Prescott, A., Lane, D. P., and Proud, C. G. Rapid induction of apoptosis mediated by peptides that bind initiation factor eIF4E. Curr. Biol., *10:* 793–796, 2000.

117. DeFatta, R. J., Li, Y., and De Benedetti, A. Selective killing of cancer cells based on translational control of a suicide gene. Cancer Gene Ther., *9:* 573–578, 2002.